Tạp chí Tin học và Điều khiển học, T.28, S.3 (2012), 260–273

TIẾP CẬN ĐẠI SỐ GIA TỬ TRONG XÂY DỰNG THANG ĐIỂM NGÔN NGỮ DÙNG TRONG ĐÁNH GIÁ

NGUYỄN CÁT HỒ, TRẦN THÁI SƠN

Viện Công nghệ thông tin, Viện Khoa học và Công nghệ Việt Nam

Tóm tắt. Bài báo trình bày phương pháp xây dựng bộ 4 ngữ nghĩa theo cách tiếp cận Đại số gia tử dùng trong đánh giá cho trường hợp thang điểm là một tập con của tập các hạng từ có độ dài không vượt quá số k cho trước.

Abstract. This paper presents methods of formulating the semantic 4-tuple approach basing on Hedge algebra used in case assessment scale which is a subset of the term set of lengths less than a given number k.

1. MỞ ĐẦU

Bài toán đánh giá các đối tượng căn cứ vào các ý kiến giáo viên (chuyên gia) là một bài toán đã được nghiên cứu từ lâu do tính chất phổ biến của nó trong nhiều lĩnh vực của đời sống như giáo dục, kinh tế, lựa chọn ra quyết định... Có thể hình thức hóa bài toán như sau: có m chuyên gia (giáo viên, hội đồng chấm tuyển..) $J_1, J_2, ..., J_m$ đánh giá n đối tượng (học sinh, nhà thầu...) $O_1, O_2, ..., O_n$ theo các tiêu chí $c_1, c_2, ..., c_k$. Các kết quả đánh giá của chuyên gia J_i căn cứ tiêu chí c_p cho đối tượng O_q ký hiệu là $x^i_{p,q}$. Yêu cầu đặt ra là xây dựng phương pháp kết nhập (aggregation) các $x^i_{p,q}$ theo q để cho ra một kết quả có giá trị đánh giá, sắp xếp các đối tượng O_q theo một thứ tự hợp lý. Trong trường hợp các $x^i_{p,q}$ là các giá trị số (thực), người ta thường sử dụng các phép lấy trung bình (cộng, nhân, có trọng số hoặc không...) làm phương pháp kết nhập. Kết quả nhận được chính là điểm tổng hợp để đánh giá các đối tượng. Vấn đề trở nên phức tạp hơn khi vì nhiều lí do, việc đánh giá không cho ra các giá trị số mà có thể là các từ ngôn ngữ tự nhiên thuộc một tập $\boldsymbol{S}$ cho trước nào đó, như "giỏi", "khá tốt", "hơi kém"... Có một số cách tiếp cận để giải quyết vấn đề này [1,3,5-9].

Phương pháp tính toán ngôn ngữ dựa trên nguyên lý mở rộng của tập mờ

Ý tưởng chính của phương pháp là các phép tính kết nhập kinh điển như phép trung bình số học, trung bình có trọng số..., có thể chuyển thành các phép tính tương ứng trên các tập mờ. Các phép kết nhập mờ thực hiện trên các tập mờ của các nhãn trong tập $\boldsymbol{S}$ sẽ cho kết quả là tập mờ. Nhược điểm dễ thấy của phương pháp sử dụng lý thuyết tập mờ trong trường hợp này là tập mờ kết quả thường là một hàm không biểu thị cho một nhãn ngôn ngữ nào trong $\boldsymbol{S}$, gây khó khăn cho việc xử lý tiếp theo.

Phương pháp tính toán trên các ký hiệu ngôn ngữ

Trong phương pháp này, ý tưởng là sắp tuyến tính các từ trong thang điểm đánh giá và dùng chỉ số thứ tự tương ứng của mỗi từ trong thang điểm thay cho từ đó để thực hiện việc kết nhập số học. Phương pháp này đơn giản về mặt tính toán nhưng chứa đựng sai số có thể là rất lớn trong trường hợp ngữ nghĩa của thang điểm đánh giá phân bố không "đều" trong miền giá trị.

Phương pháp tính toán ngôn ngữ dựa trên biểu diễn dữ liệu bộ 2

Trong phương pháp trên ta cần làm tròn kết quả kết nhập để có thể ứng kết quả với một từ ngôn ngữ trong thang điểm đánh giá. Tuy nhiên việc làm tròn làm mất mát thông tin và các tác giả [10] đã đưa ra cách biểu diễn dữ liệu bộ 2 để khắc phục sự mất mát thông tin này. Ý tưởng của phương pháp là ngoài giá trị kết nhập làm tròn ta còn lưu thêm giá trị sai số làm tròn đó để dùng trong các xử lý tiếp theo. Tuy nhiên, về thực chất, đây cũng chỉ là một cải tiến chứ không thay đổi được bản chất của phương pháp là lấy chỉ số thứ tự của từ ngôn ngữ thay cho giá trị ngữ nghĩa của từ đó. Và như đã nói, phương pháp dạng này đơn giản về mặt thực thi nhưng có thể gây sai số lớn, chưa kể nó làm mất hẳn ý nghĩa cốt lõi của việc dùng thang điểm đánh giá ngôn ngữ.

Phương pháp tính toán ngôn ngữ dựa trên biểu diễn dữ liệu bộ 3, bộ 4 [4]

Theo hướng nghiên cứu này, các tác giả đã nghiên cứu phát triển việc tính toán trên các từ dựa trên lý thuyết về đại số gia tử (ĐSGT) với việc sử dụng các khoảng độ đo tính mờ và giá trị định lượng ngữ nghĩa của các từ ngôn ngữ. Với việc sử dụng các độ đo tính mờ nói chung, việc xử lý các từ ngôn ngữ được gắn chặt với ngữ nghĩa, đặc biệt là quan hệ thứ tự tự nhiên của chúng. Do đó, việc mất mát thông tin được hạn chế tối đa. Đồng thời, quá trình xử lý cũng như kết quả thu được là dễ dàng cảm nhận được theo tư duy con người. Trong các nghiên cứu này, tập các giá trị đánh giá (thang điểm) được xét bao gồm tất cả các hạng từ của ĐSGT có độ dài cố định hoặc không vượt quá một số cho trước. Trong bài báo này sẽ mở rộng theo hướng xét thang điểm linh hoạt hơn, là một *tập con của tập tất cả hạng từ có độ dài không vượt quá một số cho trước*. Việc này tuy dẫn đến các phương pháp xử lý phức tạp hơn nhưng phù hợp tự nhiên vì trong thực tế, phần lớn các trường hợp người ta cho thang điểm không đầy đủ tất cả các từ nêu ở trên. Khi đó việc tính toán không đơn thuần là chỉ bỏ qua các giá trị không có trong thang điểm mà phải xây dựng lại toàn bộ miền giá trị của thang điểm. Thí dụ, nếu thang điểm chỉ có 3 từ {"giỏi", "khá", "kém"} thì "khá" ở đây, trong tương quan với các hạng từ còn lại, khác với "khá" trong thang điểm {"giỏi", "khá", "tương đối khá", "kém"}. Người ta nói ở đây hạng từ có tính chất phụ thuộc ngữ cảnh.

Mục 2 sẽ nhắc lại một số kiến thức cần thiết về ĐSGT sẽ được sử dụng trong bài báo. Phương pháp hình thức hóa biểu diễn ngữ nghĩa bằng bộ 4 các từ ngôn ngữ trong thang đánh giá sẽ được trình bày trong Mục 3. Các phép kết nhập trên dữ liệu biểu diễn bộ 4 cũng được định nghĩa trong mục này. Mục 4 sẽ mô tả các cách giải bài toán lấy quyết định bằng các phương pháp khác nhau và bằng phương pháp mới được đề xuất trong bài báo. Việc phân tích so sánh các kết quả của các phương pháp trong ví dụ minh họa ở Mục 5 sẽ chứng tỏ lợi ích và hiệu quả của phương pháp mới.

2. MỘT SỐ KHÁI NIỆM CƠ BẢN

Định nghĩa 2.1. [11] Đại số gia tử được ký hiệu là $AX = (X, G, H, \leq)$ là bộ bốn, trong đó X là miền giá trị của biến ngôn ngữ, G là tập các phần tử sinh và hằng, H là tập các gia tử (hedge) xét như các toán tử một ngôi tác động lên các hạng từ thuộc X, còn $\leq$ là quan hệ cảm sinh ngữ nghĩa trên X. Giả thiết trong G có chứa các phần tử là hằng số $0, 1, W$ với ý nghĩa là phần tử bé nhất, phần tử lớn nhất và phần tử trung bình (neutral) trong X. Ta gọi mỗi giá trị ngôn ngữ $x \in X$ là một hạng từ (term) trong ĐSGT.

Nếu tập X và H là các tập sắp thứ tự tuyến tính, khi đó $AX = (X, G, H, \leq)$ là ĐSGT tuyến tính. Ngoài ra, thông thường trong các ứng dụng, miền giá trị của biến ngôn ngữ bao gồm các từ sinh ra từ hai phần tử sinh đối xứng (như "trẻ" và "già", "xa" và "gần", "cao" và "thấp"...) nên trong bài, kể từ đây nói ĐSGT cũng có nghĩa là ĐSGT tuyến tính có 2 phần tử sinh đối xứng, ký hiệu là c^+ và c^-. Như vậy, $G = \{0, c^-, W, c^+, 1\}$.

Thí dụ ĐSGT có X là miền giá trị của biến ngôn ngữ "Tuổi", là tập các hạng từ như {"rất già", "già", "rất rất trẻ", "tương đối trẻ", "tương đối rất trẻ"...}, với $G = \{0, \text{Trẻ}, W, \text{Già}, 1\}$ và H ={'rất", "tương đối",..} có quan hệ $\leq$ cảm sinh ngữ nghĩa như "rất già">"già">"tương đối trẻ">"tương đối rất trẻ" ">"rất rất trẻ"...

Khi tác động gia tử $h \in H$ vào phần tử $x \in X$, thì thu được phần tử ký hiệu hx. Với mỗi $x \in X$, ký hiệu $H(x)$ là tập tất cả các hạng từ $u \in X$ sinh từ x bằng cách áp dụng các gia tử trong H và viết $u = h_n...h_1x$, với $h_n, ..., h_1 \in H$, n sẽ là độ dài của u, ký hiệu $l(u)$. Tập H gồm các gia tử dương H^+ và gia tử âm H^-. Các gia tử dương làm tăng ngữ nghĩa của một hạng từ mà nó tác động, còn gia tử âm làm giảm ngữ nghĩa của hạng từ. Trong bài báo này, ta giới hạn việc nghiên cứu ở ĐSGT hai gia tử, ký hiệu tương ứng là $V \in H^+$ và $L \in H^-$. Để biểu diễn tính dương, âm của một gia tử đối với một hạng từ, ta có hàm dấu Sign. Nếu gia tử h dương đối với x thì $\text{Sign}(hx) = 1$, ngược lại, $\text{Sign}(hx) = -1$.

Cũng như lý thuyết tập mờ, một khái niệm mờ phải được lượng hóa (thông qua hàm thuộc –membership functions), để giải quyết các bài toán thực tế, trong ĐSGT, các giá trị ngôn ngữ cũng có thể lượng hóa. Điểm khác biệt là nhờ các tính chất của ĐSGT, việc lượng hóa này bảo toàn được quan hệ thứ tự có sẵn giữa các hạng từ và cho phép trong khi tiến hành một số phép biến đổi, vẫn giữ được ngữ nghĩa của các hạng từ cùng kết quả kết nhập – là một việc hầu như không thể đối với việc sử dụng các hàm thuộc theo Zadeh. Việc lượng hóa các hạng từ của ĐSGT được tiến hành thông qua khái niệm tính mờ của các giá trị ngôn ngữ. Dưới góc độ của ĐSGT - "tính mờ của một hạng từ x được hiểu như là ngữ nghĩa của nó vẫn có thể được thay đổi khi tác động vào nó bằng các gia tử" [11]. Vậy tập các hạng từ sinh từ x nhờ tác động các gia tử sẽ thể hiện tính mờ của x và do đó $H(x)$ có thể sử dụng như là một mô hình biểu thị tính mờ của x và kích thước tập $H(x)$ được xem như độ đo tính mờ của x. Ta có thể xây dựng hàm độ đo tính mờ fm như một ánh xạ từ X vào [0,1], ánh xạ tập $H(x)$ vào khoảng $fm(x)$, thỏa mãn một số tính chất được coi như tiên đề. Tiếp theo, ta xây dựng các khoảng $\Im(x)$ có độ dài bằng độ dài của $fm(x)$ và nằm trên đoạn [0,1] (là miền giá trị chuẩn hóa của X) theo thứ tự tương ứng với thứ tự của x, thì theo tính chất của ĐSGT, với tập tất cả các hạng từ x_i có độ dài k cho trước, ta có một phân hoạch đoạn [0,1] thành các khoảng tính mờ $\Im(x_i)$. Tiếp theo, trong mỗi khoảng tính mờ $\Im(x_i)$ có thể chọn một điểm đại diện cho cả giá trị ngôn ngữ x_i, mà trong ĐSGT gọi là giá trị định lượng ngữ nghĩa (ĐLNN) $\Im(x_i)$ theo một công thức xác định thông qua một thủ tục tính đệ quy, xuất phát từ những tham số cho trước của ĐSGT là $fm(c^-), fm(c^+), \mu(h_j), h_j \in H$, ở đó $\mu(h_j)$ là độ đo tính mờ của gia tử h_j. Cũng theo tính chất của ĐSGT dễ thấy giá trị định lượng ngữ nghĩa $\Im(x)$ của hạng từ x chính là điểm cuối chung của hai khoảng tính mờ $\Im(Lx)$ và $\Im(Vx)$. Giá trị này chia khoảng tính mờ $\Im(x)$ theo tỷ lệ $\alpha : \beta$ nếu $\mathrm{Sign}(Vx) = 1$, hoặc tỷ lệ $\beta : \alpha$ nếu $\mathrm{Sign}(Vx) = -1$, trong đó α, β tương ứng là độ đo tính mờ của c^- và c^+. (xem [2]).

Các định nghĩa hình thức và tính chất của ĐSGT có thể xem ở [5,11].

Trước khi đi vào nội dung chính của bài báo, ta khảo sát một số tính chất về láng giềng của các khoảng tính mờ để làm cơ sở cho việc xây dựng thang điểm ngôn ngữ. Ta đưa vào các ký hiệu thường dùng sau:

$\boldsymbol{X}_k$ là tập các hạng từ có độ dài đúng bằng k,

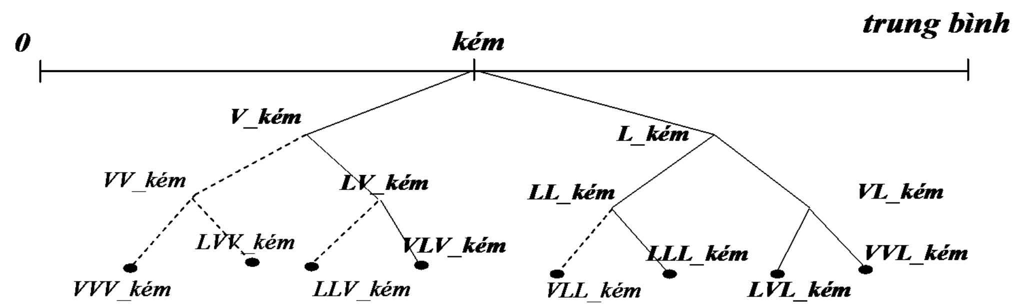
$\boldsymbol{X}_{(k)} = \cup_{l=1,\dots,k}\boldsymbol{X}_l$ là tập tất cả các hạng từ có độ dài từ 1 đến k và rõ ràng $\boldsymbol{X} = \bigcup_{k=1}^{\infty} \boldsymbol{X}_k$, và

$\boldsymbol{I}_k = \{\Im_k(x) : x \in \boldsymbol{X}_k\}$ là tập tất cả các khoảng tính mờ bậc k,

$\boldsymbol{I} = \{\Im(x) : x \in X\} = \bigcup_{k=1}^{\infty} \boldsymbol{I}_k$.

Tương tự ta cũng có tập $\boldsymbol{I}(k) = \bigcup_{l=1,\dots,k}^{\infty} \boldsymbol{I}_l$.

Để trình bày đơn giản, ta sử dụng biểu diễn cây nhị phân của tập thang điểm S các giá trị ngôn ngữ. Ta nhớ lại rằng đường đi trong một đồ thị T (không hướng) là một dãy các đỉnh liên tiếp $x_0, x_1, ..., x_k$ sao cho (x_{i-1}, x_i) là một cạnh của đồ thị T, và chỉ xuất hiện duy nhất một lần trong dãy này, $i = 1, ..., k$. Cây là một đồ thị liên thông không có chu trình và có một đỉnh được lấy làm gốc.



Hình 2.1. Biểu diễn cây của tập giá trị đánh giá $T(c^-) \subseteq X_{(4)}$

Từ tập các hạng từ được sắp tuyến tính, $\mathbf{0} < c < \boldsymbol{W} < c^+ < \mathbf{1}$, ta xây dựng một đồ thị T gồm các đỉnh đơn $\mathbf{0}, \boldsymbol{W}$ và $\mathbf{1}$ và hai cây có gốc là hai phần tử sinh c^+ và c^- có độ cao không vượt quá l cho trước, được mô tả như sau. Ta kí hiệu đồ thị con của T sinh ra từ hạng từ $c \in \{c^-, c^+\}$ là $T(c)$. Tại mỗi nút có nhãn là hạng từ x của ĐSGT, $x \in H(c)$, ta có 2 cung đi đến 2 đỉnh là có nhãn là các hạng từ Vx và Lx: hạng từ nằm ở nhánh trái có thứ tự nhỏ hơn hạng từ nằm nhánh phải. Chẳng hạn, trên hình vẽ 1.1, ta có $V_{\text{kém}} < L_{\text{kém}}$ hay $LVL_{\text{kém}} < VVL_{\text{kém}}$, theo tính chất của ĐSGT, nên, theo định nghĩa của T các đỉnh $V_{\text{kém}}$ và $LVL_{\text{kém}}$ nằm ở các nhánh trái tương ứng so với các đỉnh $L_{\text{kém}}$ và $VVL_{\text{kém}}$. Trên Hình 1.1, các nút nối bằng các cung biểu diễn bằng đường liền là các hạng từ thuộc $T(c)$, còn các nút đi đến bằng cung biểu diễn bằng đường gạch là các hạng từ không thuộc $T(c)$ (để cho tiện, từ đây ta đồng nhất tập thang điểm S với đồ thị T có các nút có nhãn thuộc S, nghĩa là ta nói nút (hay đỉnh) $x \in T$, nghĩa là nút có nhãn $x \in S$). Ở đây, ta giả thiết tập giá trị ngôn ngữ của thang điểm T thuộc tập $\boldsymbol{X}_{(l)}$, với l là số tự nhiên cho trước, và thỏa mãn tính chất: $\forall x \in T(c)$ *thì mọi đỉnh trên đường đi từ gốc c của cây* $T(c)$ *đến* x *cũng đều thuộc* $T(c)$. (*)

Khi đó, từ cách xây dựng cây $T(c)$ ta thấy $hx \in T(c) \Longrightarrow x \in T(c)$ và ta nói $T(c)$ là đóng đối với hạng từ sinh trực tiếp hay, dùng ngôn ngữ đồ thị, $T(c)$ có tính chất (*). Tính chất này phản ánh đòi hỏi tự nhiên là nếu một hạng từ dạng hx thuộc thang điểm thì hạng từ x sinh trực tiếp ra hx cũng thuộc thang điểm. Ví dụ nếu hạng từ "rất khá" thuộc thang điểm T thì hạng từ "khá" cũng phải thuộc T, nếu không, bản thân từ "rất khá" sẽ mang nghĩa không đầy đủ khi hạng từ "khá" không xuất hiện cùng với "rất khá" trong T. Tính chất này đảm bảo rằng đồ thị con $T(c)$ là liên thông và rõ ràng nó không có chu trình, nghĩa là nó là một cây nhị phân. Ở đây, ta luôn giả thiết T có tính chất đóng với sinh trực tiếp (*).

3. MỘT SỐ TÍNH CHẤT CỦA TẬP THANG ĐIỂM NGÔN NGỮ T

Vì tập T các hạng từ ngôn ngữ luôn sắp tuyến tính, để dễ hình dung ta quy ước là với mọi hạng từ $x, y \in T,\ x < y$ khi và chỉ khi x nằm ở dải bên trái (chiếu theo hoành độ trong Hình 1.1) của y và y nằm dải bên phải của x trong đồ thị T.

Với quy ước đó, dựa theo tính chất của ĐSGT, đồ thị T có các tính chất sau:

- $T(c^-)$ nằm ở dải nằm giữa $\mathbf{0}$ và $\boldsymbol{W}$, còn $T(c^+)$ nằm ở dải giữa $\boldsymbol{W}$ và $\mathbf{1}$.
- Với $x = h_m...h_1u \in T$, với $u \in T(c)$, dãy các hạng từ $u, h_1u, h_2h_1u, ..., h_m...h_1u = x$ là một đường đi trong $T(c)$ từ gốc c đến x và đường đi này là duy nhất đi từ u đến x (vì nếu trái lại T sẽ có chu trình).
- Với $x, hx \in T$ thỏa $x \leq hx (x \geq hx)$, mọi hạng từ trong $T \cap H(hx)$ đều nằm trong dải nằm bên phải của x (nằm trong dải bên trái của x).
- Với $x, hx, khx \in T$ thỏa $x \leq khx \leq hx (x \geq khx \geq hx)$, mọi hạng từ trong $T \cap H(khx)$ đều nằm trong dải nằm giữa x và hx.
- Với mỗi đỉnh x của cây, toàn bộ các đỉnh của cây con của T có gốc là đỉnh x là một tập con của tập $H(x), |x| \leq l$. Do đó, theo tính chất của ĐSGT, cho 2 cây con bất kỳ T' và T'', hoặc cây này là cây con của cây kia, hoặc mọi nút của cây con này đều nhỏ hơn

toàn bộ các nút của cây con kia. Chẳng hạn, trong Hình 1.1, cây con có gốc là $L_{\text{kém}}$ bao hàm cây con có gốc là $VL_{\text{kém}}$, còn các đỉnh của cây con có gốc $LL_{\text{kém}}$ đều nhỏ hơn các đỉnh của cây con có gốc là $VL_{\text{kém}}$.

Ưu điểm của cấu trúc ĐSGT là nó cho phép khai thác ngữ nghĩa ngữ cảnh của các hạng từ, nghĩa là một hạng từ mờ bất kì sẽ thay đổi ngữ nghĩa khi các hạng từ xuất hiện bên cạnh nó trong quan hệ thứ tự thay đổi. Vì vậy các hạng từ láng giềng có vai trò quan trọng trong nghiên cứu này. Vì thang điểm ngôn ngữ T luôn luôn giả thiết là sắp tuyến tính nên mỗi hạng từ x của T có 2 láng giềng trong T, một láng giềng trái, kí hiệu là x_L, một láng giềng phải, kí hiệu là x_R, trừ hạng từ đầu tiên của T chỉ có láng giềng phải và hạng từ cuối cùng của T chỉ có láng giềng trái. Nhớ rằng trong ĐSGT x có dạng xâu, $x = h_{n-1}...h_2h_1c$, với độ dài $|x| = n$. Với $0 < p < n$, xâu con $h_{p-1}...h_2h_1c$ của x được gọi là khúc đầu thứ p của x và kí hiệu bằng $x|_p$. Khi đó, $x|_1 = c$.

Bổ đề 3.1. *Nếu láng giềng trái x_L (phải x_R) của x trong $T(c)$ thỏa $x, x_L \in T(c)$ $(x, x_R \in T(c)), c \in \{c^-, c^+\}$, thì tồn tại duy nhất một đường đi từ gốc c đến một trong hai hạng từ x và x_L sao cho hạng từ còn lại cũng nằm trên đường đi. Trường hợp x_L là nút cuối đường đi, ta có $x = x_L|_{|x|}$, còn trường hợp x là nút cuối đường đi, ta có $x_L = x|_{|x_L|}$.*

Chứng minh. Vì $x, x_L \in T(c)$ và, theo tính chất (T2) nêu ở trên, tồn tại một đường đi duy nhất từ gốc c đến $x_L, c, h_1c, h_2h_1c, ..., h_q...h_2h_1c = x_L$. Mặt khác, do $T(c)$ là một cây nên nó liên thông. Vì vậy, tồn tại một đường đi nối hai nút x_L và x đi qua lần lượt các nút $x_L = u_0, u_1, ..., u_{p-1}, u_p = x$. Khi đó, theo cách định nghĩa của cây $T(c)$, hoặc nó đi ra từ nút u_0, nghĩa là ta có $u_1 = k_1x_L$ và đường đi có dạng $x_L, k_1x_L, k_2k_1x_L, ..., k_pk_{p-1}...k_1x_L = x$, hoặc nó đi ra từ nút u_p, nghĩa là $u_{p-1} = k_1x$ và khi đó đường đi có dạng $x, k_1x, k_2k_1x, ..., k_pk_{p-1}...k_1x = x_L$. Trong trường hợp thứ nhất ta thu được đường đi duy nhất từ gốc c đến x_L đi qua nút x và, do đó, $k_i = h_i, i = 1, ..., p < q$, hay ta có $x = x_L|_{p+1}$. Tương tự, trong trường hợp thứ hai ta thu được đường đi duy nhất từ gốc c đến x đi qua nút x_L và, do đó, $x_L = x|_{q+1}$.

Tương tự, ta có chứng minh cho trường hợp của x_R. ■

Bổ đề 3.2. *Với $x \in X_{(l)}$ thỏa $|x| = l$ thì láng giềng trái x_L (láng giềng phải x_R) của x trong $X_{(l)}$ phải thỏa $|x_L| < l(|x_R| < l)$.*

Chứng minh. Tính đúng đắn của bổ đề dễ dàng suy ra từ Bổ đề 3.1. ■

Định lý 3.1. *Giả sử $T \subseteq X_{(l)}$ là tập các giá trị đánh giá thỏa mãn tính chất (*) $(l > 1)$ và $x = h_m...h_1c \in T,\ c \in \{c^-, c^+\}$. Nếu $x_L = h'q...h'1c$ là láng giềng trái của x trong T thì*

(L1) *Với $q < m$, ta có*

(i) $x_L = x|_q = h_q...h_1c$ *và* (ii) $x_L < h_{q+1}x_L$.

(L2) *Với $q > m$, ta có*

(i) $x_L = h'_q...h'_{m+1}x$ *và* (ii) $h'_{m+1}x < x$.

Nếu $x_R = h'_q...h'_1c$ là láng giềng phải của x trong T, thì

(R1) *Với $q < m$, ta có*

(i) $x_R = x|_q = h_q...h_1c$ *và* (ii) $h_{q+1}x_R < x_R$.

(R2) *Với $q > m$, ta có*

(i) $x_R = h'_q ... h'_{m+1}c$ *và* (ii) $x < h'_{m+1}x$.

Chứng minh. Ta chứng minh (L1).

Vì $x, x_L \in T(c)$ và do $p < m$, theo Bổ đề 3.1, tồn tại một đường đi duy nhất từ c đến x đi qua x_L và $x_L = x|_q$. Vì $x_L < x$, nghĩa là x nằm bên phải của x_L, nên theo tính chất (T3) áp dụng đối với x_L và $h_{q+1}x_L$ với lưu ý rằng $x \in H(h_{q+1}x_L) \cap T$, các hạng từ trong $H(h_{q+1}x_L) \cap T$ cần phải nằm trong dải bên phải của x_L. Vậy, $x_L < h_{q+1}x_L$.

Lập luận tương tự, với $q > m$, từ Bổ đề 3.1 ta có một đường đi đến x_L đi qua $x, x = x_L|_{m+1}$ và $h'_{m+1}x < x$, nghĩa là (L2) được chứng minh.

Chứng minh tương tự với thay đổi đối ngẫu giữa $<$ và $>$ ta có thể chỉ ra tính đúng của (R1) và (R2) cho hạng từ lân cận phải x_R của x. ■

Định lí 3.2. *Cũng với T như trên, với $x \in T$, nếu $x_L(x_R)$ là láng giềng trái (phải) của x trong T nhưng không phải trong $X_{(l)}$ thì $p_L = \max(|x_L|, |x|) < l$ ($p_R = \max(|x_R|, |x|) < l$) và nó cũng là láng giềng trái (phải) của x trong $X_{(p_L)}(X_{(pR)})$. Trong trường hợp riêng $l = 1$, vì có phần tử x của $X_{(l)}$ nằm giữa x và x_L (hoặc x_R), nên $x_L(x_R)$ sẽ là láng giềng trái (phải) của x trong $X_{(l)}$.*

Chứng minh. Giả sử x và x_L không cùng nằm trong một cây. Đầu tiên ta xét trường hợp $x \in T(c)$ và khi đó, $x_L \in \{\mathbf{0}, \boldsymbol{W}\}$, tập các hạng từ gắn với cây chỉ có một đỉnh trong đồ thị T. Giả thiết rằng tồn tại láng giềng trái $x'_L \neq x_L$ của x trong $X_{(p_L)}$. Vì $X_{(p_L)}$ cũng là một cây, nên tồn tại hai đường đi D và D', D đi từ gốc c đến x và D' đi từ c đến x'_L. Giả sử x^* nút cuối của đường đi chung của hai đường D và D'. Khi đó, Vx^* và Lx^* sẽ là hai nút không nằm trên cùng một đường đi, nghĩa là một nằm trên D, chẳng hạn đó là Vx^*, còn hạng từ Lx^* nằm trên D' và chúng nằm hai phía khác nhau của nút x^*. Theo (T5), đường đi con còn lại của D và D' tương ứng cùng với các hạng từ Vx^* và Lx^* nằm bên phải hoặc bên trái của x^*. Vì $x'_L < x$, nên ta suy ra $Lx^* < Vx^*$ và khi đó ta có $x'_L < x^* < x$. Vì theo tính chất (*), $x^* \in T$, ta suy ra x_L không phải là láng giềng trái của x trong T, xuất hiện mâu thuẫn. Vậy x_L cũng phải là láng giềng trái của x trong $X_{(p_L)}$. Trường hợp $x_L \in T(c)$ (khi đó $x \in \{\boldsymbol{W}, \mathbf{1}\}$) chứng minh hoàn toàn tương tự.

Bây giờ giả sử $x, x_L, x_R \in T(c)$. Vì chứng minh cho trường hợp x_R hoàn toàn tương tự, nên ta chứng minh cho trường hợp láng giềng trái và giả sử $|x| = \max(|x_L|, |x|)$. Vì $|x_L| < |x|$, theo Bổ đề 3.1, có một đường đi duy nhất D từ c đến x đi qua x_L và $x_L = x|_{|x_L|}$. Giả sử tồn tại $x'_L \neq x_L$ là láng giềng trái của x trong $X_{(p_L)}$. Khi đó $x'_L \notin T$. Vì $X_{(l)} \cap H(c)$ cũng là một cây và, theo Bổ đề 3.2, $|x'_L| < |x|$, nên lại theo Bổ đề 3.1, tồn tại trong cây này một đường đi duy nhất D' từ c đến x đi qua x'_L. Vì $D \notin D'$ trong cây $X_{(l)} \cap H(c)$, nên chúng sẽ chứa một chu trình, mâu thuẫn với định nghĩa của cây. Mâu thuẫn thu được chứng tỏ x_L cũng là láng giềng trái của x trong $X_{(p_L)}$.

Bây giờ ta giả sử $|x_L| = p_L > |x|$ và $x'_L \neq x_L$ cũng là láng giềng trái của x trong $X_{(p_L)}$. Gọi D và D' tương ứng là hai đường đi duy nhất từ c đến x_L và từ c đến x'_L. Gọi x^* là nút cuối của đường đi chung của D và D'. Bằng lập luận tương tự như phần đầu của chứng minh định lí ta thu được hoặc $x_L < x^* < x'_L < x$ hoặc $x'_L < x^* < x_L < x$. Vì theo tính chất (*), $x^* \in T$, ta suy ra x_L hoặc x'_L không phải là láng giềng trái của x tương ứng trong T hoặc trong $X_{(p_L)}$. Cả hai trường hợp đều dẫn đến mâu thuẫn. Vậy x_L cũng phải là láng giềng trái của x trong $X_{(p_L)}$. Định lý hoàn toàn được chứng minh. ■

4. XÂY DỰNG THANG ĐIỂM NGÔN NGỮ

Bây giờ ta đi xây dựng thang điểm ngôn ngữ, cụ thể hơn là xây dựng một hệ khoảng lân cận cho các hạng từ thuộc T của ĐSGT AX, lấy hệ này làm cơ sở để thiết lập các tham số dành cho tính toán trong phép kết nhập các hạng từ thuộc thang điểm đã cho. Ta biết, với một số k cho trước, ta có một phân hoạch các khoảng tính mờ trên miền giá trị (chuẩn hóa) $[\mathbf{0}, \mathbf{1}]$, ở đó mỗi khoảng tính mờ $\Im_k(x)$ tương ứng với một hạng từ x có độ dài k của ĐSGT. Do nhu cầu thực tế, việc chỉ xét đến các hạng từ có cùng độ dài là không đủ, như trong bài toán đánh giá nêu ở đây chẳng hạn, chúng ta phải xét đến các hạng từ có độ dài khác nhau của ĐSGT, cụ thể là tập các hạng từ $x \in X_{(k)}$ là tập các hạng từ có độ dài không lớn hơn k. Khi đó, ta cũng có thể xây dựng một hệ khoảng lân cận bậc k ứng với các hạng từ $x \in X_{(k)}$ để rồi thiết lập cơ chế tính toán dựa trên các khoảng này (xem [4,11]), gọi là khoảng k-tương tự $\{\mathcal{S}_k(x)|_x \in X_{(k)}\}$, thỏa mãn 2 điều kiện sau:

(i) Các khoảng k-tương tự này tạo nên một phân hoạch trên miền giá trị chuẩn hóa [0,1].

(ii) Mỗi $\mathcal{S}_k(x)$ chứa một và chỉ một giá trị định lượng ngôn ngữ $\nu(x)$ và các giá trị thuộc $\mathcal{S}_k(x)$ được coi là k-tương đương với $\nu(x)$.

Trong trường hợp cụ thể ta đang xét, khi tập H chỉ gồm 2 phần tử V và L, có thể xác định tường minh như sau:

$$\mathcal{S}_k(x) = [\nu(x_{(L,k+1)}), \nu(x_{(k,k+1)})) = cup\{\Im(y) : |y| = k+2, \Im(y) \subseteq [\nu(x_{(L,k+1)}), \nu(x))$$

hoặc

$$\Im(y) \subseteq [\nu(x), \nu(x_{(R,k+1)}))\}. \quad (1)$$

Tiếp theo, cũng xuất phát từ thực tế, ta thấy thang điểm dùng đánh giá trong nhiều trường hợp không bao gồm tất cả các hạng từ thuộc $X_{(k)}$ mà chỉ một tập con S của nó. Do S chỉ là tập con của $X_{(k)}$ nên không thể dùng khoảng k-tương tự này cho S (vì sẽ có những khoảng lân cận không tham gia, phá vỡ tính chất phân hoạch của các khoảng lân cận). Tư tưởng là ta phải kết nối một số khoảng lân cận của $X_{(k)}$ thành một hệ khoảng lân cận mới dùng cho S, vì thực tế, ngữ nghĩa của các giá trị đánh giá thuộc S đã thay đổi so với ngữ nghĩa ban đầu của từ này trong $X_{(k)}$. Nói cách khác, ngữ nghĩa của một từ là phụ thuộc ngữ cảnh, nó thay đổi tùy theo sự xuất hiện trong những tập khác nhau mà ta xét (tập từ chỉ giá trị đánh giá hay tập các hạng từ của ĐSGT có độ dài không lớn hơn k). Thí dụ nếu ta chỉ dùng hai từ "tốt" và "xấu" để đánh giá một đối tượng thì phạm vi ngữ nghĩa của từ "tốt" là khá rộng, trong khi nếu ta thêm vào tập đánh giá các từ "rất tốt" và "khá tốt" chẳng hạn thì ngữ nghĩa của từ "tốt" đã bị thu hẹp lại đáng kể. Với tư tưởng ấy, sử dụng các kết quả về láng giềng mục trên, chúng ta sẽ xây dựng khoảng lân cận cho các hạng từ của S, gọi là khoảng ngữ nghĩa, với hạn chế là mỗi tập gia tử H^+ và H^- chỉ có đúng một phần tử (là V và L).

Khoảng ngữ nghĩa của x trong S, nửa đoạn $I(x) = [a, b)$ được xác định dựa trên đoạn trái con $I_L(x) = [a, \nu(x)]$ và nửa đoạn phải con $I_R(x) = [\nu(x), b)$. Ký hiệu khoảng k-tương tự $\mathcal{S}_k(x)$ của x trong $X_{(k)}$ theo mức k là $[\text{lep}\mathcal{S}_k(x)_k, \text{rep}\mathcal{S}_k(x)) = [\text{lep}\mathcal{S}_k(x), \nu(x)] \cup [\nu(x), \text{rep}_k\mathcal{S}_k(x))$, ở đó $\text{lep}\mathcal{S}_k(x), \text{rep}\mathcal{S}_k(x)$ tương ứng lần lượt là điểm cuối trái (left end-point) và phải (right end-point) của khoảng k-tương tự của x trong $X_{(k)}$.

Định nghĩa 4.1. Cho ĐSGT AX. Khi đó, với mỗi $x \in S$, khoảng ngữ nghĩa của x trong S được xác định là khoảng $I(x) = I_L(x) \cup I_R(x)$, ở đó $I_L(x) = L\mathcal{S}_{PL}(x) = [lpt\mathcal{S}_{PL}(x), \nu(x)]$ với $p_L = \max(|x_L|, |x|)$ và $I_R(x) = R\mathcal{S}_{PR}(x) = [\nu(x), rpt\mathcal{S}_{PR}(x))$ với $p_R = \max(|x_R|, |x|)$. Rõ ràng, giá trị định lượng ngữ nghĩa $\nu(x) \in I(x)$ và bậc tương tự của các phần tử của $I(x)$ là $p = \min(p_L, p_R)$.

Tính chất 4.1. *Tập các khoảng ngữ nghĩa $I(x)$ xác định trong Định nghĩa 3.1 tạo ra một phân hoạch trong [0,1]. Hơn nữa, nếu cả hai phần tử x_L, x_R là láng giềng trái và phải của x trong $X_{(l)}$ thì $I(x) = \mathcal{S}_k(x)$, với $k = p_L = p_R$.*

Chứng minh. Theo định nghĩa láng giềng trái phải và cách xây dựng $I(x)$ và cả Tính chất 2.3 dễ dàng suy ra được điều phải chứng minh.

Như vậy mỗi phần tử $x \in T$ của ĐSGT AX có thể ứng với khoảng $I(x)$ là một khoảng ngữ nghĩa của x. Điều đó dẫn đến định nghĩa sau.

Định nghĩa 4.2. Cho ĐSGT AX, với mỗi $x \in S$ của AX, bộ bốn biểu diễn ngữ nghĩa của x là bộ bốn có dạng $(x, I(x), \nu(x), r_x)$, trong đó $I(x)$ là khoảng ngữ nghĩa của $x, \nu(x)$ là giá trị định lượng ngữ nghĩa của x và $r_x \in I(x)$ là một giá trị số (do ta gắn hoặc là kết quả của một phép tính nhất định). Ngoài ra, tập $\{I(x) : x \in S\}$ tạo nên một phân hoạch đoạn $[0, 1]$ và $\forall x, y \in S, x \leq y$ dẫn đến $I(x) \leq I(y)$.

Bộ bốn ngữ nghĩa nêu trong định nghĩa trên có thể gán giá trị số hoặc không, tùy vào trường hợp cụ thể, khi chuyên gia đánh giá bằng điểm hoặc bằng hạng từ ngôn ngữ hoặc bằng số, như đã nói trong phần mở đầu. Trong trường hợp x được gán giá trị số, ta có tính chất sau.

Tính chất 4.2. *Cho bộ bốn ngữ nghĩa $(x, I(x), \nu(x), r_x)$, mỗi phép gán giá trị số r_x được biểu diễn bởi một bộ bốn duy nhất, được ký hiệu là $(x(r_x), I(x(r_x)), \nu((x(r_x)), r_x)$. Nếu $r_x \leq r'_x$ thì các hạng từ $x(r_x)$ và $x(r'_x)$ xuất hiện trong bộ bốn biểu diễn các giá trị số r_x, r'_x thỏa mãn bất đẳng thức $x(r_x) \leq x(r'_x)$.*

Chứng minh. Vì các khoảng ngữ nghĩa $I_{\partial(s)}(s), s \in S$ tạo nên một phân hoạch của đoạn [0,1] nên với mỗi $r_x \in [0, 1]$, tồn tại duy nhất một hạng từ s sao cho $r_x \in I_{\partial(s)}(s)$. Ngoài ra, khoảng ngữ nghĩa và giá trị ĐLNN của nó cũng xác định duy nhất do đó bộ bốn biểu diễn cũng là xác định duy nhất. Tiếp theo, dựa vào Định nghĩa 3.2, ta có $r_x \leq r'_x$ dẫn đến $I(x_x) \leq I(r'_x)$ và suy ra $x(r_x) \leq x(r'_x)$. ■

Định nghĩa 4.3. Cho ĐSGT $AX = (\boldsymbol{X}, \boldsymbol{G}, \boldsymbol{H}, \leq)$. Tập các từ $S = \{s_0, ..., s_g\} \subseteq X$ được gọi là thang điểm ngôn ngữ với mức l đóng với hạng từ sinh trực tiếp hay có tính chất (*) nếu thỏa mãn các điều kiện sau:

(i) $\exists s \in S$ sao cho $|s| = l$ và S có thứ tự toàn phần theo $\leq$.

(ii) $c^-, c^+, \mathbf{0}, \boldsymbol{W}, \mathbf{1} \in S$.

(iii) S có tính chất (*).

Định nghĩa 4.4. Thang điểm ngữ nghĩa bộ bốn mức l của thang điểm ngôn ngữ S là tập các bộ bốn $S_{Q.l} = \{(s, I_{\partial(s)}(s), Q(s), r_s) : s \in S, r_s \in [0, 1]\}$ thỏa mãn các điều kiện sau, ở đây [0,1] là chuẩn hóa của miền tham chiếu của biến ngôn ngữ và ∂_s chỉ bậc tương tự của các giá trị của $I_{\partial(s)}(s)$ với ngữ nghĩa của s:

(i) S là tập các từ bất kỳ được sắp thứ tự toàn phần dựa trên ngữ nghĩa của chúng.

(ii) $I_{\partial(s)}(s), s \in S$, là khoảng ngữ nghĩa của s và chúng tạo ra một phân hoạch của đoạn [0,1]. Ngoài ra, $s \in S' \Longrightarrow I_{\partial(s)}(s) \leq I_{\partial(s')}(s')$.

(iii) $Q(s)$ là ngữ nghĩa định lượng bằng số của s và cả $Q(s), r_s \in I_{\partial(s)}(s)$.

Thang điểm định nghĩa như vậy có nhiều ưu việt và mang nhiều thông tin phục vụ cho việc đánh giá các đối tượng hay phương án trong môi trường mờ. Đầu tiên, khác biệt với thang điểm dựa trên tập mờ, ở đây chúng được xây dựng dựa trên ngữ nghĩa định tính của các từ ngôn ngữ dựa trên một cơ sở hình thức hóa chặt chẽ. Nó mô phỏng thang điểm ngôn ngữ trong thực tế mà các khoảng điểm ứng với các từ ngôn ngữ được xác định theo kinh nghiệm của người sử dụng. Nhưng trong thang điểm bộ bốn có thành phần $Q(s)$ mang thông tin giống như tâm của tập mờ: nó là giá trị số phù hợp với ngữ nghĩa của từ ngôn ngữ s nhất. Còn thành phần điểm số r có hai ý nghĩa thực tiễn. Một là nó biểu thị đánh giá của chuyên gia bằng điểm số trong bối cảnh có thang điểm ngôn ngữ, tạo sự thống nhất trong sử dụng hai thang điểm đồng thời. Hai là, nó bảo đảm tính đóng của phép kết nhập, một phép tính rất quan trọng trong việc đánh giá chuyên gia đối với các bài toán quyết định đa tiêu chuẩn đối với các phương án lựa chọn.

Định lí 4.1. *Cho ĐSGT $AX = (X, G, H, \leq)$ và tập các từ $S = \{s_0, ..., s_g\} \subseteq X$ là thang điểm ngôn ngữ có tính chất (*) với mức l dựa trên nền AX. Khi đó mỗi bộ tham số của ĐSGT AX xác định một bộ bốn thang điểm ngữ nghĩa với mức $lSQ.l = \{(s, \nu(s), r(s), I_{\partial(s)}(s)) : s \in S, r(s) \in [0, 1]\}$, thỏa mãn các tính chất:*

(i) Mỗi khoảng $I_{\partial(s)}(s)$ được xác định dựa trên ngữ nghĩa của hạng từ của $X : I_{\partial(s)}(s) = I_L(s) \cup I_R(s)$ và đấy là hợp của các khoảng mờ liên tiếp của các hạng từ xác định của X_{l+2}.

(ii) Với mỗi $s \in S$ và $r_s \in I_{\partial(s)}(s), (s, \nu(s), r(s), I_{\partial(s)}(s))$ là biểu diễn ngữ nghĩa ngôn ngữ bộ bốn của s.

(iii) $r_s \leq r'_s$ suy ra $s \leq s'$.

Chứng minh.

(i) Theo Định nghĩa 3.2, khoảng tương tự mức l được xác định duy nhất theo tập các tham số của ĐSGT AX. Vì S có tính chất (*) ở mức l nên ta có $S \subseteq X_{(l)}$ và $\forall s \in S$, láng giềng trái và phải của s được xác định theo Tính chất 2.3. Khi đó khoảng ngữ nghĩa $I(s)$ được xác định theo Định nghĩa 3.2, ở đó $I(s) = I_L(s) \cup I_R(s)$. Vì mỗi khoảng tương tự mức l là hợp của một số các khoảng mờ của các hạng từ trong X_{l+2} nên với $k < l+2$, mỗi khoảng k-mờ là hợp của các khoảng $(l+2)$-mờ xác định, từ đó theo Định nghĩa 3.1 suy ra $I(s)$ là hợp của các khoảng $(l+2)$- mờ xác định thuộc X_{l+2}.

(ii) Như đã chứng minh ở trên, $\nu(s), I_{\partial(s)}(s) = I_L(s) \cup I_R(s)$ có thể tính (tự động) theo các tham số của ĐSGT và những giá trị số của $I_{\partial(s)}(s)$ được xem là tương tự ngữ nghĩa với s, cũng là với giá trị định lượng ngữ nghĩa $\nu(s)$ với mức $\partial(s) = \min(p_L, p_R)$.

Như vậy, $I_{\partial(s)}(s)$ có thể xem như khoảng ngữ nghĩa của s. Ta biết tập $\{I_{\partial(s)}(s) : s \in S\}$ tạo nên một phân hoạch của đoạn [0,1] nên ta có điều phải chứng minh.

(iii) Theo Tính chất 3.2. ■

Định nghĩa 3.5. Giả sử cho 2 bộ bốn $(x, I(x), \nu(x), r)$ và $(x', I(x'), \nu(x'), r')$. Ta có thể viết $(x, I(x), \nu(x), r) \leq (x', I(x'), \nu(x'), r')$ khi và chỉ khi một trong các điều kiện sau thỏa mãn:

(i) $x \leq x'$

(ii) $x = x'$ và $r \leq r'$.

Theo cách xác định khoảng ngữ nghĩa cho các phần tử của thang điểm, ta có tính chất sau:

Định lí 4.2. *Cho ĐSGT $AX = (X, G, H, \leq)$ và S là tập có tính chất (*) bậc l dựa trên AX. Khi đó mỗi khoảng ngữ nghĩa $I_{\partial(s)}(s)$ được xác định và tính dựa trên ngữ nghĩa của các hạng từ của $AX : I_{\partial(s)}(s) = I_L(s) \cup I_R(s)$ và*

$$I_{\partial(s)}(s) == \bigcup\{\Im(x) : x \in X_{l+2} \& \Im(x) \subseteq [\nu(S_{L,P_L+1}), \nu(S_{R,P_R+1})\}.$$

Chứng minh. Vì S có tính chất (*) bậc l nên $S \subseteq X_{(l)}$ và với mỗi $s \in S, I_{\partial(s)}(s) = I_L(s) \cup I_R(s)$ với $\partial(s) = \min(p_L, p_R)$, được xác định theo Tính chất 3.3, tức là $I_{\partial(s)}(s)$ có thể xác định và tính dựa trên các khoảng k-tương tự và cụ thể hơn là dưa trên các khoảng tính mờ.

Theo Định nghĩa 3.1 ta có $I_L(s) = L\mathcal{S}_{P_L}(s)$ ở đó $p_L = \max(|s_L|, |s|) \leq l$ với s_L là láng giềng trái s trong S. Nếu láng giềng trái và phải của s trong X_{P_L+1} là $s_{L_{p_L}+1}$ và $s_{R_{p_L}+1}$, theo công thức (1),

$$\mathcal{S}_{P_L}(s) = [\nu(S_{L,P_L+1}), \nu(S_{R,P_L+1})) = \bigcup\{\Im(z) : z \in X_{P_L+2} \& \Im(z) \subseteq [\nu(S_{R,P_L+1}), \nu(s))$$

hoặc

$$\Im(z) \subseteq [\nu(s), \nu(S_{L,P_L+1}))\}.$$

Ta suy ra $I_L(s) = [\nu(S_{L,P_L+1})), \nu(S_{R,P_L+1})) = \bigcup\{\Im(z) : z \in X_{P_L+2} \& \Im(z) \subseteq [\nu(S_{L,P_L+1}), \nu(s))$. Vì $p_{L+2} \leq l+2$ và theo quy nạp mỗi phần tử của X_{P_L+2} là hợp của các khoảng tính mờ của X_{l+2}, ta có

$I_L(s) = \bigcup\{\Im(z) : z \in X_{l+2} \& \Im(z) \subseteq [\nu(S_{L,P_L+1}), \nu(s))\}$. Tương tự với láng giềng phải $I_R(s) = R\mathcal{S}_{P_R}(s) = [\nu(s), \nu(S_{R,P_R+1})) = \bigcup\{\Im(z) : z \in X_{l+2} \& \Im(z) \subseteq [\nu(s), \nu(S_{R,P_R+1})$. Vì $I(s) = I_L(s) \cup I_R(s)$, ta có điều phải chứng minh. ■

Định nghĩa 4.6. Cho thang điểm ngữ nghĩa bộ bốn với mức l của thang điểm ngôn ngữ $S, S_{Q.l}$, và phép toán kết nhập số α theo nghĩa của Yager trong đoạn [0,1]. Phép toán α sẽ cảm sinh phép kết nhập Agg_α trên $S_{Q.l}$ được xác định như sau:

$$Agg_\alpha(a_1, ..., a_n) = (x(r_a), I(x(r_a)), \nu(x(r_a)), r_a)$$

trong đó a là vecto của các biểu diễn bộ bốn của các hạng từ $a = (a_1, ..., a_n), a_i = (x_i, I(x_i), \nu(x_i), r_i),\ i = 1, ..., n,\ r_a = \alpha(r_1, ..., r_n)$ và $(x(r_a), I(x(r_a)), \nu(x(r_a)), r_a)$ là biểu diễn bộ bốn xác định duy nhất bởi r_a.

Định lí 4.3. *Thang điểm ngữ nghĩa ngôn ngữ bộ bốn được xây dựng có tính chất đóng đối với phép kết nhập Agg_α. Hơn thế, Agg_α là đơn điệu tăng và nếu α là đơn điệu tăng chặt thì Agg_α cũng thế.*

Chứng minh. Theo định nghĩa Agg_α dễ thấy tính chất đóng thỏa mãn. Cũng theo định nghĩa của từng thành phần, dễ thấy tính đơn điệu tăng của Agg_α. Bây giờ, giả sử $\boldsymbol{a} = (a_1, ..., a_n) < \boldsymbol{b} = (b_1, ..., b_n)$, ta suy ra $\exists i = 1..n : a_i = (x_i, I(x_i), \nu(x_i), r_{a,i}) < b_i = (y_i, I(y_i), \nu(y_i), r_{b,i})$. Từ Định nghĩa 3.5, ta có $r_{a,i} < r_{b,i}$. Theo tính chất đơn điệu chặt của α, ta có $r_a = \alpha(r_{a,1}, ..., r_{a,n}) < r_b = \alpha(r_{b,1}, ..., r_{b,n})$, dẫn đến $(t(r_a), I(t(r_a), \nu(t(r_a)), r_a) < (t(r_b), I(t(r_b), \nu(t(r_b)), r_a)$, vì $t(r_a) \leq t(r_b)$. ■

5. VÍ DỤ MINH HỌA

Xét một thang điểm bao gồm các giá trị đánh giá S = {"hoàn toàn kém", "rất kém", "kém", "tương đối kém", "trung bình", "khá", "giỏi", "rất giỏi", "xuất sắc"}. Căn cứ vào ngữ nghĩa, có thể tạo ra thang điểm như vậy từ ĐSGT AX gồm 2 phần tử sinh "kém" và "giỏi" cùng hai gia tử "rất" và "tương đối". Ở đó **0** ứng với "hoàn toàn kém", $\boldsymbol{w}$ ứng với "trung bình", "tương đối giỏi" ứng với "khá"..., **1** ứng với "xuất sắc". Để cho gọn, chúng tôi sử dụng ký hiệu thay cho các từ thuộc ĐSGT như sau: hai phần tử sinh "kém" và "giỏi" tương ứng là c^- và c^+; hai gia tử "tương đối" và "rất" tương ứng là L và V. Thang điểm S khi đó sẽ là $S = \{0, Vc^-, c^-, Lc^-, w, Lc^+, c^+, Vc^+, 1\}$. Ta cũng giả sử có 1 chuyên gia đánh giá 2 đối tượng $A1$ và $A2$ dựa trên S theo 2 tiêu chí nào đó và cho kết quả $A1 = \{Lc^-, Vc^+\}, A2 = \{w, c^+\}$. Nếu theo phương pháp bộ 2 (đã nói trong phần mở đầu) dễ thấy kết quả tổng hợp đánh giá của 2 đối tượng này là như nhau (cùng bằng 6 – "khá", nếu ứng chỉ số cho các phần tử thuộc S theo thứ tự từ 1 đến 9 và dùng phép kết nhập là trung bình cộng). Theo cách tiếp cận ĐSGT, có thể thấy các bộ 4 tương ứng sẽ là:

$s_1 = (Lc^-, I(Lc^-), \nu(Lc^-), r_1) : r_1 \in I(Lc^-); s_2 = (Vc^+, I(Vc^+), \nu(Vc^+), r_2) : r_2 \in I(Vc^+)$;

$s_3 = (w, I(w), \nu(w), r_3) : r_1 \in I(w); s_4 = (c^+, I(c^+), \nu(c^+), r_4) : r_4 \in I(c^+)$;

Thực hiện phép gán các giá trị thuộc ĐSGT lấy các giá trị ĐLNN tương ứng và tiến hành các phép lấy trung bình cộng trên các giá trị này ta sẽ có:

$T_1 = [\nu(Lc^-) + \nu(Vc^+)]/2 = \{[\mu^3(V) + 3\mu^2(V)\mu(L) + 2\mu(V)\mu^2(L)]fm(c^-) + fm(c^-) + [\mu^3(L) + 2\mu^2(V)\mu(L) + 3\mu(V)\mu^2(L)]fm(c^+)\}/2.$

$T_2 = [\nu(w) + \nu(c^+)]/2 = \{2fm(c^-) + \mu(L)fm(c^+)\}/2.$

Để so sánh T_1 và T_2, thực hiện một số biến đổi với lưu ý rằng $fm(c^-) + fm(c^+) = \mu(L) + \mu(V) = 1$, ta có $2(T_1 - T_2) = \mu(L)[\mu(V)fm(c^+) - \mu(L)fm(c^-)]$.

Từ đó suy ra $T_1 = T_2 \iff \mu(V)fm(c^+) = \mu(L)fm(c^-); T_1 > T_2 \iff \mu(V)fm(c^+) > \mu(L)fm(c^-)$.

Chúng ta có nhận xét, trong trường hợp sử dụng phương pháp bộ 2, tức là thực hiện phép so sánh dựa trên phép kết nhập theo chỉ số thứ tự của giá trị đánh giá trong thang điểm, thì dù có quan niệm về việc bố trí các giá trị đánh giá có thay đổi thế nào, kết quả so sánh vẫn không thay đổi, tức là hai đối tượng được đánh giá như nhau. Ngược lại, theo cách tiếp cận ĐSGT, phụ thuộc vào việc bố trí, kết quả đánh giá có thay đổi. Chẳng hạn, nếu giả thiết $\mu(V) = \mu(L)$, tức là nếu coi tác động của gia tử lên các giá trị biến ngôn ngữ là như nhau (chỉ ngược hướng), thì việc so sánh T_1 với T_2 tương đương với việc so sánh $fm(c^-)$ và $fm(c^+)$. Cụ thể, khi $fm(c^-) > fm(c^+)$, tức là khi thang điểm được hình dung là "khắt khe" (phần dành cho đánh giá hơn trung bình ít hơn phần dành cho đánh giá dưới trung bình) thì $T_2 > T_1$, tức là ("Trung bình"+"Giỏi") > ("Tương đối kém"+"Rất giỏi"), hay nói cách khác, do thang điểm kiểu "khắt khe" nên đánh giá "Tương đối kém" sẽ được coi là nặng, kéo theo đánh giá kết nhập sẽ thấp hơn so với đánh giá ("Trung bình"+"Giỏi"). Điều tương tự ngược lại cũng xảy ra khi thang điểm được coi là kiểu "dễ dãi" , tức là $fm(c^-) < fm(c^+)$. Dấu "=" xy xảy ra khi $fm(c^-) = fm(c^+)$, tức là khi các giá trị ngôn ngữ của thang điểm phân bố đều trên miền xác định. Ta thấy qua ví dụ, cách tiếp cận ĐSGT có vẻ hợp lý hơn về mặt tổng thể.

6. KẾT LUẬN

Cách tiếp cận bài toán kết nhập theo ĐSGT cho ta một kết quả hợp lý hơn so với các cách tiếp cận dựa vào chỉ số thứ tự. Trong bài báo này, chúng tôi tiếp tục nghiên cứu xây dựng bộ bốn dùng trong xử lý toán tử kết nhập dành cho trường hợp thang điểm không đầy đủ các phần tử của một ĐSGT. Mặc dù vậy, như trong [3] đã nhận xét, có những trường hợp dù theo cách tiếp cận nào cũng gây sai số và làm nẩy sinh các vấn đề không mong muốn (như "không dân chủ - vì bỏ qua hoàn toàn ý kiến của một số chuyên gia). Cách duy nhất khắc phục phần nào là sử dụng thang điểm sinh thêm (thêm vào những giá trị đánh giá có độ dài hạng từ lớn hơn).

TÀI LIỆU THAM KHẢO

[1] Nguyễn Cát Hồ, Trần Thái Sơn, Về khoảng cách giữa các giá trị của biến ngôn ngữ trong Đại số gia tử, *Tạp chí Tin học và Điều khiển học* **11** (1) (1995) 10–20.

[2] Nguyễn Cát Hồ, Trần Thái Sơn, Dương Thăng Long, Tiếp cận Đại số gia tử cho phân lớp mờ, *Tạp chí Tin học và Điều khiển học* **25** (1) (2009) 53–68.

[3] Trần Thái Sơn, Nguyễn Tuấn Anh, Bài toán kết nhập mờ theo cách tiếp cận Đại số gia tử, *Tạp chí Khoa học và Công nghệ, Đại học Thái Nguyên* **81** (5) (2011) 97–101.

[4] Nguyễn Văn Long, Hoàng Văn Thông, Vấn đề kết nhập thông tin biểu diễn bằng bộ 4 với ngữ nghĩa dựa trên Đại số gia tử, *Tạp chí Tin học và Điều khiển học* **27** (3) (2011).

[5] Nguyễn Văn Long, Cơ sở phương pháp luận của phương pháp đánh giá bằng nhãn ngôn ngữ, *Tạp chí Tin học và Điều khiển học* **24** (1) (2008) 75–86.

[6] M. Delgado, F. Herrera, E. Herrera-Viedma, L. Marffnez, Combining numerical and linguistic information in group decision making, *Journal of Information Sciences* **107** (1998) 177–194.

[7] F. Herreraa, E. Herrera-Viedmaa, Luis Martnez, A fusion approach for managing multi-granularity linguistic term sets in decision making, *Fuzzy Sets and Systems* **114** (2000) 43–58.

[8] F. Herrera, E. Herrera-Viedma, Linguistic decision analysis: steps for solving decision problems under linguistic information, *Fuzzy Sets and Systems* **115** (2000) 67–82.

[9] F. Herrera and L. Martinez, A 2-Tuple Fuzzy Linguistic Reoresentation Model for Computing with Words, *IEEE Transactions on Systems, Man, and Cybernetics* **8** (6) (2000) 746–752.

[10] F. Herrera and L. Martinez, A model based on linguistic 2-tuples for dealing with multigranular hierachical linguistic contexts in multi-expert decision-making, *IEEE Transactions on Systems, Man, and Cybernetics* **31** (2) (2001) 227–234.

[11] N.C. Ho, N.V. Long, Fuzziness measure on complete hedge algebras and quantitative semantics of terms in linear hedge algebras, *Fuzzy Sets and Systems* **158** (4) (2007) 452–471.

[12] Van Hung Le, Cat Ho Nguyen, Fei Liu, Semantics and Aggregation of Linguistic Information Based on Hedge Algebras, The Third International Conference on Knowledge, *Information and Creativity Support Systems, KICSS 2008*, Hanoi, Vietnam, Dec. 22-23, 2008 (128–135).

[13] Van Nam Huynh, Cat Ho Nguyen, Yoshiteru Nakamori, MEDM in general multi-granular hierarchical linguistic contexts based on the 2-tuples linguistic model, *Proc. of IEEE Int. Conf. on Granular Computing*, 2005 (482–487).

[14] Jun Liu, Da Ruan, Roland Carchon, Synthesis and evaluation analysis of the indicator information in nuclear safeguards applications by computing withwords, Int, *J. Appl. Math. Comput. Sci.* **12** (3) (2002) 449–462.

Ngày nhận bài 17 - 7 - 2012

Nhận lại sau sửa ngày 18 - 12 - 2012