Tạp chí Tin học và Điều khiển học, T. 14, S. 3 (1998) (35–41)

ỨNG DỤNG ĐẠI SỐ GIA TỬ ĐỐI SÁNH CÁC GIÁ TRỊ NGÔN NGỮ

TRẦN ĐÌNH KHANG

Abstract. This paper interpretes some methods for comparison of linguistic values. There are the aggregation of hedge algebras, the if-then rules and the distance table between linguistic values. It can be used for linguistic reasoning in practical problems.

1. ĐẶT VẤN ĐỀ

Ngày nay, lý thuyết tập mờ và suy luận xấp xỉ được ứng dụng rộng rãi trong các hệ chuyên gia, hệ hỗ trợ quyết định và trong điều khiển, tạo ra khả năng xử lý các giá trị "không rõ", trong đó có một lớp quan trọng là các giá trị ngôn ngữ, ví dụ như "Anh ấy tương đối cao". Chính lớp các giá trị ngôn ngữ này đóng vai trò rất quan trọng trong nhận thức và tư duy của con người. Một cách tư duy tương đối thông dụng là suy luận ví dụ như chúng biết rằng: "Một người khoẻ thì chạy nhanh", thì ta cũng có thể đánh giá được khả năng chạy của một đối tượng A, khi "A rất khoẻ". Kết luận sẽ phụ thuộc nhiều vào sự đối sánh giữa "khỏe" và "rất khỏe" để cho giá trị tương ứng với "nhanh".

Trong lý thuyết tập mờ, tập mờ được xây dựng dựa vào hàm thuộc trên tập vũ trụ làm giá đỡ cho tập mờ đó, từ đó phần nào tạo ra khả năng đối sánh giữa các tập mờ nhờ tập vũ trụ như một yếu tố để "quy đồng mẫu số". Tuy nhiên, việc định nghĩa tập vũ trụ cho một biến mờ, định nghĩa hàm thuộc cho một tập mờ vẫn có thể chứa đựng nhiều yếu tố không chính xác. Và bản thân luật mờ kéo theo (implication) do được mô hình hóa như là phép kéo theo của logic đa trị, trong khi tồn tại quá nhiều toán tử kéo theo khác nhau, tạo ra khả năng có các mô hình khác nhau cho một quá trình suy diễn.

Một cách giải quyết là thao tác trực tiếp trên các giá trị ngôn ngữ, một trong các mô hình được nghiên cứu và phát triển trong những năm gần đây là đại số gia tử ([1]) đặt các giá trị ngôn ngữ vào một cấu trúc đại số gồm các phần tử sinh và các gia tử (từ nhấn), từ đó có thể so sánh được giữa các giá trị này. Tuy nhiên, một biến ngôn ngữ cũng có thể chứa đựng các giá trị ngôn ngữ hết sức đa dạng, khó có thể đặt chung vào một cấu trúc đại số gia tử, như nói về sốt, có thể có {sốt nặng, sốt nhẹ, sốt vừa, sốt theo cơn, sốt kiên tục, sốt về chiều, sốt kèm gai rét, sốt đổ mồ hôi,...}.

Trong bài này chúng tôi sẽ đưa ra các phương pháp để đối sánh các giá trị ngôn ngữ dựa trên cấu trúc đại số gia tử, hàm đo trên đại số gia tử ([2]) và tích hợp đại số gia tử ([3]), làm cơ sở cho suy luận ngôn ngữ. Các phương pháp này được minh họa bằng hệ chẩn đoán bệnh đang được nghiên cứu và thử nghiệm tại Phòng nghiên cứu ứng dụng các hệ chuyên gia, hệ hỗ trợ quyết định thuộc Viện Công nghệ thông tin.

2. ỨNG DỤNG ĐẠI SỐ GIA TỬ ĐỂ ĐỐI SÁNH GIỮA CÁC GIÁ TRỊ NGÔN NGỮ

Ta quan tâm đến các biến ngôn ngữ có các giá trị đa dạng không dễ dàng đưa về một cấu trúc đại số gia tử như biến sốt ở trên. Rõ ràng việc so sánh các giá trị trong biến này là hoàn toàn không thể, nếu không có thêm thông tin cần thiết. Bây giờ, đặt tập các giá trị trên vào mục đích chẩn đoán bệnh viêm gan, với giá trị đặc trưng sốt của viêm gan là "sốt nhẹ, không thành cơn, trong khoảng 1 tuần đến 10 ngày, kèm theo nhức đầu,...", từ đó tạo ra một cấu trúc đại số gia tử tổng hợp sốt_của_viêm_gan. Khi đó tập các giá trị của sốt có thể xếp được vào trong cấu trúc này, việc so sánh chúng trở nên có thể, và thậm chí tập giá trị còn trở nên phong phú hơn nhờ việc tác động của các từ nhấn vào các phần tử trong cấu trúc.

Vấn đề đặt ra là cách xây dựng đại số gia tử tổng hợp, hay chính xác hơn là các mô hình để các chuyên gia - bác sĩ có thể cung cấp thông tin (tri thức) giúp cho việc tạo ra đại số gia tử tổng hợp.

2.1. Tích hợp đại số gia tử

Từ tập giá trị của một biến ngôn ngữ, có thể đưa về nhiều đại số gia tử khác nhau, như với biến sốt:

- mức độ sốt: sốt nhẹ, sốt vừa, sốt nặng,... (X_1)
- kiểu sốt: thành cơn, liên tục, thành nhiều đợt,... (X_2)
- thời gian sốt: sáng, chiều, đêm,... (X_3)
- số ngày sốt: 1 ngày, 2 ngày, ngắn, sốt dài ngày,... (X_4)
- kèm nhức đầu: nhức đầu, không nhức đầu,... (X_5)
- kèm rét run: rét, không rét run,... (X_6)
- kèm đổ mồ hôi: đổ mồ hôi, không đổ mồ hôi,... (X_7)

Đại số gia tử sốt_của_viêm_gan được xây dựng dựa trên việc tích hợp các đại số gia tử $X_1, X_2, ..., X_7$. Cách tích hợp thông dụng là theo trọng số, có một tập trọng số $\{p_1, p_2, ..., p_7\}$ với tổng các p_i bằng 1, thể hiện mức độ ngữ nghĩa của các đại số gia tử thành phần tham gia vào đại số gia tử chung. Trong [3] đã trình bày định nghĩa hàm tích hợp đại số gia tử, xin nêu lại dưới đây.

Định nghĩa. Cho $\chi \subseteq X_1 \times X_2 \times \cdots \times X_n$ là bộ n đại số gia tử với các trọng số $p_1, p_2, ..., p_n$; hàm đo λ và hàm khoảng cách θ. X^* là đại số gia tử tổng hợp. Ta có $f_w : \chi \to X^*$ là hàm tích hợp đại số gia tử tại $w \in \chi$ xác định đầy đủ cho trước, nếu $\forall x, y \in \chi$ xác định đầy đủ, sao cho $\theta(x, w) \leq \theta(y, w)$ thì

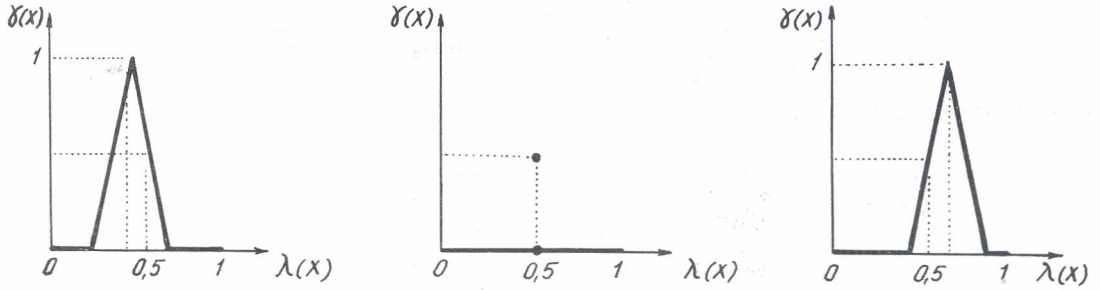
$$|\lambda(f_w(x)) - \lambda(f_w(w))| \leq |\lambda(f_w(y)) - \lambda(f_w(w))|.$$

Như vậy $w = (w_1, w_2, ..., w_n)$ chính là giá trị ngôn ngữ đặc trưng. $x = (x_1, x_2, ..., x_n) \in \chi$ là xác định đầy đủ nếu các thành phần x_i đều được xác định (có giá trị). Vấn đề cần giải quyết là xây dựng một dạng hàm thỏa mãn định nghĩa trên, áp dụng cho đại số gia tử tổng hợp. Để làm được điều này, ta định nghĩa các *hàm tương tự thành phần* $\gamma_i : X_i \to [0, 1]$ chỉ ra mức độ giống nhau giữa các giá trị ngôn ngữ đưa và giá trị ngôn ngữ đặc trưng thuộc thành phần X_i. Cho $x_i, y_i \in X_i$, hàm γ_i cần phải thỏa mãn các điều kiện sau:

- $\gamma_i(w_i) = 1$
- γ_i (Unknow) = 0,5
- Nếu $|\lambda(x_i) - \lambda(w_i)| \leq |\lambda(y_i) - \lambda(w_i)|$ thì $\gamma_i(x_i) \geq \gamma_i(y_i)$

Sau đây là một hàm tương tự thỏa mãn các điều kiện trên

$$\gamma_i = \begin{cases} 1 - \dfrac{|\lambda(x) - \lambda(w_i)|}{2\,|0,5 - \lambda(w_i)|} & \text{với } |\lambda(x) - \lambda(w_i)| < 2\,|0,5 - \lambda(w_i)| \\ 0,5 & \text{với } 0 = |\lambda(x) - \lambda(w_i)| = 2\,|0,5 - \lambda(w_i)| \\ 0 & \text{với } 0 \neq |\lambda(x) - \lambda(w_i)| \geq 2\,|0,5 - \lambda(w_i)| \end{cases}$$



a/ $\lambda(w_i) < 0,5$ b/ $\lambda(w_i) = 0,5$ c/ $\lambda(w_i) > 0,5$

Ví dụ: Giá trị đặc trưng sốt của viêm gan là "sốt nhẹ, không thành cơn, trong khoảng 1 tuần đến 10 ngày, kèm theo nhức đầu,...", thì

$\gamma_{\text{mức độ sốt}}$ (sốt nhẹ) = 1,

$\gamma_{\text{kèm nhức đầu}}$ (sốt ít nhức đầu) = 0,6.

...

Rõ ràng, hàm tương tự sẽ cho giá trị lớn nếu khoảng cách giữa các giá trị ngôn ngữ đưa vào và giá trị đặc trưng thành phần là nhỏ và ngược lại. Nếu giá trị đưa vào là Unknow thì hàm tương tự cho giá trị 0,5.

Khi đó, cho $\underline{x} = \{x_1, x_2, ..., x_n\}$ với $x_i \in X_i$, ta có thể tính được

$$\lambda(\underline{x}) = \sum_{i=1}^{n} p_i * \gamma_i(x_i)\,. \qquad (*)$$

Tiếp theo, dùng hàm ngược của hàm đo, tính được $\underline{x}$ thuộc đại số gia tử tổng hợp.

Định lý. *Cho $\chi \subseteq X_1 \times X_2 \times \cdots \times X_n$ là bộ của n đại số gia tử với các trọng số $p_1, p_2, ..., p_n$; hàm đo λ và hàm khoảng cách θ. Các hàm tương tự thành phần được định nghĩa như công thức trên với $w = (w_1, w_2, ..., w_n)$ xác định đầy đủ là giá trị ngôn ngữ đặc trưng. Khi đó, hàm*

$$f_w(x) = \lambda^{-1}\Big(\sum_{i=1}^{n} p_i * \gamma_i(x_i)\Big)$$

là hàm tích hợp đại số gia tử tại w.

Chứng minh. Cho $x = (x_1, x_2, ..., x_n)$, $y = (y_1, y_2, ..., y_n) \in \chi$.

Đặt
$$I_{x_1} = \{i,\ 1 \le i \le n,\ |\lambda(x_i) - \lambda(w_i)| < 2\,|0,5 - \lambda(w_i)|\}$$
$$I_{x_2} = \{i,\ 1 \le i \le n,\ 0 = |\lambda(x_i) - \lambda(w_i)| = 2\,|0,5 - \lambda(w_i)|\}$$
$$I_{x_3} = \{i,\ 1 \le i \le n,\ 0 \ne |\lambda(x_i) - \lambda(w_i)| \ge 2\,|0,5 - \lambda(w_i)|\}$$
$$I_{y_1} = \{i,\ 1 \le i \le n,\ |\lambda(y_i) - \lambda(w_i)| < 2\,|0,5 - \lambda(w_i)|\}$$
$$I_{y_2} = \{i,\ 1 \le i \le n,\ 0 = |\lambda(y_i) - \lambda(w_i)| = 2\,|0,5 - \lambda(w_i)|\}$$
$$I_{y_3} = \{i,\ 1 \le i \le n,\ 0 \ne |\lambda(y_i) - \lambda(w_i)| \ge 2\,|0,5 - \lambda(w_i)|\}$$

Rõ ràng $I_{x_1} \cup I_{x_2} \cup I_{x_3} = I_{y_1} \cup Y_{y_2} \cup I_{y_3} = \{1, 2, ..., n\}$.

Chọn dạng hàm khoảng cách

$$\theta(x, w) = \sum_{i \in I_{x_1}} p_i \frac{|\lambda(x_i) - \lambda(w_i)|}{2\,|0,5 - \lambda(w_i)|} + \sum_{i \in I_{x_2}} 0,5\,p_i + \sum_{i \in I_{x_3}} p_i\,.$$

Khi đó với $\theta(x, w) \le \theta(y, w)$ thì có

$$\sum_{i \in I_{x_1}} p_i \frac{|\lambda(x_i) - \lambda(w_i)|}{2\,|0,5 - \lambda(w_i)|} + \sum_{i \in I_{x_2}} 0,5\,p_i + \sum_{i \in I_{x_3}} p_i$$
$$\le \sum_{i \in I_{y_1}} p_i \frac{|\lambda(y_i) - \lambda(w_i)|}{2\,|0,5 - \lambda(w_i)|} + \sum_{i \in I_{y_2}} 0,5\,p_i + \sum_{i \in I_{y_3}} p_i$$
$$\sum_{i=1}^{n} p_i - \sum_{i \in I_{x_1}} p_i \frac{|\lambda(x_i) - \lambda(w_i)|}{2\,|0,5 - \lambda(w_i)|} - \sum_{i \in I_{x_2}} 0,5\,p_i - \sum_{i \in I_{x_3}} p_i$$
$$\ge \sum_{i=1}^{n} p_i - \sum_{i \in I_{y_1}} p_i \frac{|\lambda(y_i) - \lambda(w_i)|}{2\,|0,5 - \lambda(w_i)|} - \sum_{i \in I_{y_2}} 0,5\,p_i - \sum_{i \in I_{y_3}} p_i$$
$$\sum_{i \in I_{x_1}} p_i \Big(1 - \frac{|\lambda(x_i) - \lambda(w_i)|}{2\,|0,5 - \lambda(w_i)|}\Big) + \sum_{i \in I_{x_2}} p_i(1 - 0,5) + \sum_{i \in I_{x_3}} p_i(1 - 1)$$
$$\ge \sum_{i \in I_{y_1}} p_i \Big(1 - \frac{|\lambda(y_i) - \lambda(w_i)|}{2\,|0,5 - \lambda(w_i)|}\Big) + \sum_{i \in I_{y_2}} p_i(1 - 0,5) + \sum_{i \in I_{y_3}} p_i(1 - 1)$$

Theo định nghĩa các hàm tương tự thành phần, biến đổi thành

$$\sum_{i=1}^{n} p_i * \gamma_i(x_i) \ge \sum_{i=1}^{n} p_i * \gamma_i(y_i).$$

Theo tính chất hàm ngược của hàm đo:

$$\lambda^{-1}\Big(\sum_{i=1}^{n} p_i * \gamma_i(x_i)\Big) \ge \lambda^{-1}\Big(\sum_{i=1}^{n} p_i * \gamma_i(y_i)\Big).$$

Đó chính là: $f_w(x) \ge f_w(y)$.

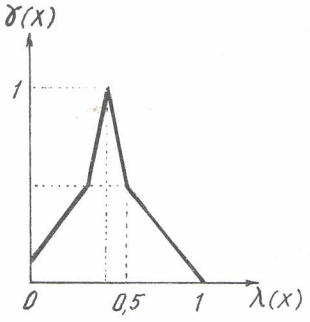
Theo định nghĩa hàm đo: $\lambda(f_w(x)) \ge \lambda(f_w(y)) \Rightarrow \lambda(f_w(x)) - 1 \ge \lambda(f_w(y)) - 1$.

Vì cả hai vế đều không dương nên: $|\lambda(f_w(x)) - 1| \le |\lambda(f_w(y)) - 1|$.

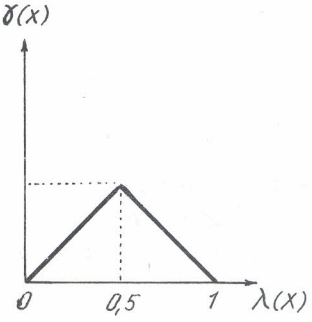
Lưu ý rằng $\lambda(f_w(w)) = \lambda\left(\lambda^{-1}\left(\sum_{i=1}^{n} p_i\right)\right) = 1$ trong trường hợp w xác định đầy đủ (các $w_i \neq$ Unknow) nên thay vào bất đẳng thức trên thỏa mãn f_w là hàm tích hợp đại số gia tử. □

Ngoài ra có thể lựa chọn các dạng khác của hàm tương tự như

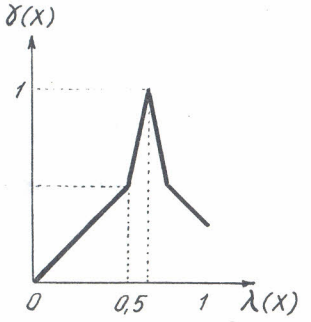
$$\gamma_i = \begin{cases} 1 - \dfrac{|\lambda(x) - \lambda(w_i)|}{2\,|0.5 - \lambda(w_i)|} & \text{với } |\lambda(x) - \lambda(w_i)| < |0,5 - \lambda(w_i)| \\ 0,5 + |0,5 - \lambda(w_i)| - |\lambda(x) - \lambda(w_i)| & \text{với } |\lambda(x) - \lambda(w_i)| \geq |0,5 - \lambda(w_i)| \end{cases}$$



a/ $\lambda(w_i) < 0,5$



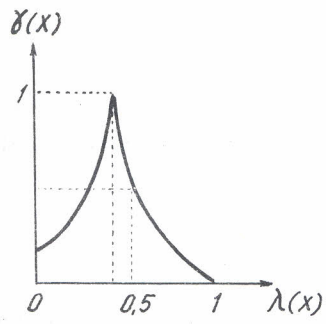
b/ $\lambda(w_i) = 0,5$



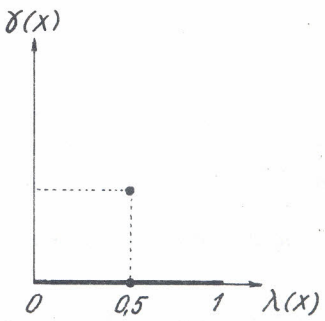
c/ $\lambda(w_i) > 0,5$

hoặc

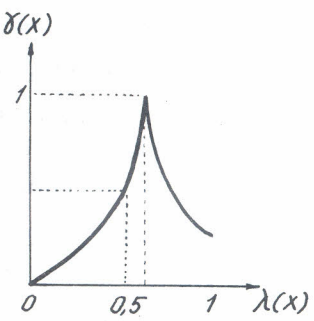
$$\gamma_i(x) = \begin{cases} \dfrac{0,5 - |0,5 - \lambda(w_i)|}{|0,5 - \lambda(w_i)|\,(0,5 + |0,5 - \lambda(w_i)|)}\,|\lambda(x) - \lambda(w_i)|^2 \\ + \dfrac{\lambda(w_i)^2 - \lambda(w_i) - |0,5 - \lambda(w_i)|}{|0,5 - \lambda(w_i)|\,(0,5 + |0,5 - \lambda(w_i)|)}\,|\lambda(x) - \lambda(w_i)| + 1 & \text{với } \lambda(w_i) \neq 0,5 \\ 0,5 & \text{với } \lambda(w_i) = 0,5 = \lambda(x) \\ 0 & \text{với } \lambda(w_i) = 0,5 \neq \lambda(x) \end{cases}$$



a/ $\lambda(w_i) < 0,5$



b/ $\lambda(w_i) = 0,5$



c/ $\lambda(w_i) > 0,5$

Đây là mô hình tổng quát cho đại số gia tử tổng hợp, tuy nhiên việc lấy tri thức sẽ rất khó khăn. Các chuyên gia có thể thống nhất với nhau trong việc phân tập giá trị thành các đại số gia tử thành phần, nhưng khó có thể đưa ra tập các trọng số p_i một cách nhất quán. Sai số của phương pháp có thể sẽ lớn.

2.2. Sử dụng các luật if-then trên các đại số gia tử thành phần

Phương pháp này dựa trên các luật kéo theo, đã được khảo sát trong [3]. Ví dụ như, từ các luật

Luật 1:	Nếu	mức độ sốt = "sốt nhẹ"
	và	kiểu sốt = "không thành cơn"
	và	số ngày sốt = "7 - 10 ngày"
	và	kèm nhức đầu = "nhức đầu"
	Thì	sốt của viêm gan = "rất đặc trưng"
Luật 2:	Nếu	mức độ sốt = "sốt vừa"
	và	kiểu sốt = "không thành cơn"
	và	số ngày sốt = "7 - 10 ngày"
	và	kèm nhức đầu = "nhức đầu"
	Thì	sốt của viêm gan = "ít đặc trưng"
Luật 3:	Nếu	mức độ sốt = "sốt cao"
	và	kiểu sốt = "thành cơn"
	Thì	sốt của viêm gan = "rất không đặc trưng"

Có thể tính được giá trị của đại số gia tử tổng hợp sốt_của_viêm_gan từ các giá trị đã biết bằng các phương pháp suy diễn mờ và suy luận xấp xỉ sẵn có dựa trên các đại số gia tử thành phần, có tập giá trị xếp vào một cấu trúc đại số gia tử.

Mô hình này có vẻ đơn giản hơn trong việc thu thập tri thức chuyên gia, nhưng lại thiếu tính mềm dẻo, đòi hỏi dữ liệu đưa vào luôn phải được phân lớp trong các đại số gia tử thành phần.

2.3. Lập bảng khoảng cách giữa các giá trị của biến ngôn ngữ

Một cách làm khác là để các chuyên gia tự đánh giá khoảng cách giữa các giá trị trong tập giá trị của biến ngôn ngữ. Các khoảng cách hai giá trị ngôn ngữ có thể là:

- ∞, nếu không sánh được với nhau, hoặc
- một giá trị số trong miền $[0, 1]$, hoặc
- một giá trị của đại số gia tử tạo thành bởi hai phần tử sinh "xa nhau" và "gần nhau".

Ngoài ra các chuyên gia cũng cho biết khoảng cách giữa các giá trị của biến ngôn ngữ với giá trị ngôn ngữ đặc trưng. Khoảng cách này có thể là một giá trị của đại số gia tử được tạo ra từ hai phần tử sinh "đặc trưng" và "không đặc trưng".

Ví dụ: Đặc trưng của ăn_uống bệnh viêm gan là "chán ăn, sợ mỡ,..."

Tập giá trị của ăn uống là

{chán ăn, ăn kém, ăn khá, sợ mỡ, sợ thịt, sợ trứng, ăn được, ăn vào đầy bụng khó tiêu, có rối loạn tiêu hóa, không có rối loạn tiêu hóa, có cảm giác đói,...}

Nếu ký hiệu θ là hàm khoảng cách thì

θ (chán ăn, ăn kém) = "rất gần nhau"

θ (ăn được, ăn_uống_của_bệnh_viêm_gan) = "không đặc trưng"

Nhận xét: Cho a, b, c là các giá trị bất kỳ của biến ngôn ngữ, thì sẽ có

$\theta(a,b) + \theta(b,c) \leq \theta(a,c)$.

Đó chính là bất đẳng thức tam giác theo định nghĩa hàm khoảng cách dùng để kiểm chứng lại bảng các khoảng cách.

Phương pháp này dễ được các chuyên gia chấp nhận. Ngoài ra, từ đó có thể đưa về đại số gia tử tích hợp theo các bước sau:

(1) Dùng các thuật toán phân lớp (cây phân cấp, k-trung bình, ISODATA,...) để phân tập giá trị ra thành n lớp (tương ứng với n đại số gia tử thành phần).

2) Gọi w là giá trị ngôn ngữ đặc trưng, x_k là các giá trị ngôn ngữ. Tính

$$h_i = \frac{1}{\text{số phần tử của lớp } i} \sum_{x_k \in \text{lớp } i} \theta(x_k, w), \quad i = 1, ..., n.$$

Nếu h_i càng nhỏ thì trọng số p_i càng lớn, cho nên có thể tính các p_i theo công thức

$$p_i = (1 - h_i) \Big/ \sum_{i=1}^{n} (1 - h_i) \quad \text{hoặc} \quad p_i = \frac{1}{h_i} \Big/ \sum_{i=1}^{n} \frac{1}{h_i}.$$

(3) Giá trị đặc trưng w bao gồm n thành phần $(w_1, w_2, ..., w_n)$, trong đó các w_i cũng thuộc tập các giá trị chung. Do vậy, có thể tính được các hàm tương tự γ_i trong lớp thứ i.

3. KẾT LUẬN

Trên đây là các phương pháp có thể cài đặt để đối sánh các giá trị ngôn ngữ của một biến ngôn ngữ theo mục đích ứng dụng cho trước. Các phương pháp trên cùng với [3], [4] tạo thành một phương pháp chung ứng dụng đại số gia tử cho các hệ chuyên gia, hệ hỗ trợ quyết định, đang được thiết kế và thử nghiệm cho việc chẩn đoán bệnh nhiệt đới tại Viện Công nghệ thông tin.

TÀI LIỆU THAM KHẢO

[1] Nguyễn Cát Hồ, W. Wechler, Extented Hedge Algebras and their Application to Fuzzy Logic, *Fuzzy Sets and System* **52** (1992) 259-281.

[2] Trần Đình Khang, Xây dựng hàm đo trên đại số gia tử và ứng dụng trong lập luận ngôn ngữ, *Tạp chí Tin học và Điều khiển học* **13** (1) (1997) 16-30.

[3] Trần Đình Khang, Tích hợp đại số gia tử cho suy luận ngôn ngữ, *Tạp chí Tin học và Điều khiển học* **13** (1997).

[4] Trần Đình Khang, Nguyễn Thanh Thủy, Xây dựng cơ chế suy diễn trên cơ sở tri thức mờ, Gửi đăng Thông báo nghiên cứu khoa học của các trường đại học.

Received: August 28, 1997

Viện Công nghệ thông tin, Trung tâm KHTN và CNQG.