Tạp chí Tin học và Điều khiển học, T. 16, S. 3 (2000), 56–64

MỘT ĐỘ ĐO LỰA CHỌN THUỘC TÍNH

ĐỖ TẤN PHONG, HỒ THUẦN, HÀ QUANG THỤY

Abstract. In this article, we propose a new measure for attribute selection (R_N - measure) having closed relations to rough measure (Pawlak Z. [5]) and R - measure (Ho Tu Bao, Nguyen Trung Dung [3]). We prove that all of these three measures are confidence measures i.e. satisfy the weak monotonous axiom. So the R_N - measure is worth in the class of attribute selection measures. Some relations between these three measures are also shown.

Tóm tắt. Bài báo đề xuất một độ đo lựa chọn thuộc tính (được ký hiệu là R_N) có quan hệ gần gũi với độ đo thô (Pawlak Z. [5]) và độ đo R (Hồ Tú Bảo, Nguyễn Trung Dũng [3]); đã chỉ ra rằng cả ba độ đo này là các độ đo tin tưởng (do thỏa mãn tiên đề đơn điệu) và như vậy R_N có vị trí trong họ các độ đo lựa chọn thuộc tính. Một số mối quan hệ liên quan đến các độ đo nói trên cũng được xem xét.

1. TIÊN ĐỀ ĐƠN ĐIỆU

Theo Dubois D. và Prade H. [1], độ đo trong lập luận xấp xỉ cần thỏa mãn tiên đề đơn điệu yếu. Tính đơn điệu của độ đo có thể được trình bày như sau:

Cho Ω là tập nào đó (được gọi là tập tham chiếu) và g là một hàm không âm xác định trên các tập con của Ω ($g : 2^{\Omega} \to R$; $\forall A \subseteq \Omega$ có $g(A) \geq 0$). Độ đo g được gọi là thỏa mãn tiên đề đơn điệu yếu (trong bài báo này được gọi tắt là tiên đề đơn điệu) nếu như:

$$\forall A, B \subseteq \Omega : A \subseteq B \text{ kéo theo } g(A) \leq g(B). \tag{1}$$

Tính đơn điệu là một trong những tính chất cốt yếu mà độ đo trong lập luận xấp xỉ cần có. Ý nghĩa của nó có thể được lý giải như sau: Khi chúng ta có được nhiều thông tin hơn trong lập luận thì độ tin cậy của lập luận sẽ được tăng lên. Tiên đề này nên được kiểm chứng mỗi khi đề xuất một độ đo trong lập luận xấp xi. Độ đo thỏa mãn tiên đề đơn điệu được gọi là độ đo tin tưởng (confidence measure).

2. ĐỘ ĐO LỰA CHỌN THUỘC TÍNH

Dữ liệu được thu từ các nguồn khác nhau thường là dữ liệu thô, mối quan hệ thông tin giữa các dữ liệu đó thường là chưa biết. Dữ liệu như vậy thường được rút ra từ các cơ sở dữ liệu quan hệ và được trình bày dưới dạng bảng chữ nhật hai chiều, trong đó mỗi hàng là dữ liệu về một đối tượng, còn mỗi cột là dữ liệu liên quan đến một thuộc tính. Một trong những mối quan hệ thông tin cần được quan tâm là sự phụ thuộc thuộc tính: Có tồn tại hay không mối quan hệ giữa nhóm thuộc tính này với một nhóm thuộc tính khác và việc lượng hóa mối quan hệ đó như thế nào? Việc xác định mức phụ thuộc giữa các nhóm thuộc tính khác nhau là một trong số các vấn đề chính trong việc phân tích, phát hiện các quan hệ nhân quả nội tại trong dữ liệu của các hệ thống. Độ đo lựa chọn thuộc tính được đặt ra nhằm mục đích giải quyết các vấn đề nói trên.

Định nghĩa 1. Giả sử O là một tập các đối tượng. $E \subseteq O \times O$ là một quan hệ tương đương trên O. Hai đối tượng $o_1, o_2 \in O$ được gọi là không phân biệt được trong E nếu chúng thỏa mãn quan hệ tương đương E (hay $o_1 E o_2$).

Định nghĩa 2. Giả sử O là một tập các đối tượng, $E \subseteq O \times O$ là một quan hệ tương đương trên O, $X \subseteq O$. Khi đó các tập $E_*(X)$ và $E^*(X)$ được định nghĩa như sau:

$$E_*(X) = \{o \in O \,|\, [o]_E \subseteq X\}, \tag{2}$$
$$E^*(X) = \{o \in O \,|\, [o]_E \cap X \neq \emptyset\}, \tag{3}$$

trong đó $[o]_E$ ký hiệu lớp tương đương gồm các đối tượng không phân biệt được với o theo quan hệ tương đương E. $E_*(X)$ và $E^*(X)$ theo thứ tự được gọi là các xấp xỉ dưới và xấp xỉ trên của X.

Xấp xỉ dưới và xấp xỉ trên được xác định theo định nghĩa trên đây đưa ra một ước lượng về tập đối tượng X nhờ phân hoạch tập đối tượng qua một quan hệ tương đương. Một số nội dung liên quan đến các xấp xỉ dưới và xấp xỉ trên cũng đã được đề cập trong [2, 4, 5, 6]. Gọi Ω là tập các thuộc tính, P là tập con của Ω. P xác định một quan hệ tương đương trên tập các đối tượng O và chia O thành các lớp tương đương, mỗi lớp bao gồm mọi đối tượng có cùng giá trị trên tất cả các thuộc tính thuộc tập thuộc tính P.

Vấn đề đặt ra là hai tập con P và Q của Ω sẽ chia O thành các lớp tương đương khác nhau và khi xem xét mối quan hệ giữa các lớp tương đương theo hai cách phân hoạch đó sẽ nhận được thông tin nhân quả nào đó giữa P và Q. Các thông tin như vậy thường được biểu diễn qua các độ đo lựa chọn thuộc tính [3].

3. ĐỘ ĐO R_N

Các độ đo lựa chọn thuộc tính trong Định nghĩa 3 và Định nghĩa 4 dưới đây đã được trình bày trong [3, 5]. Để làm ví dụ diễn giải một số nội dung, chúng ta sử dụng dữ liệu ở bảng 1 (với giả thiết không có hai hàng giống nhau do các đối tượng là phân biệt nhau từng đôi một):

Bảng 1. Bảng thông tin dữ liệu thu thập

	Nhiệt_độ	Đau_đầu	Bị_cúm
E_1	Bình_thường	Có	Không
E_2	Cao	Có	Có
E_3	Rất_cao	Có	Có
E_4	Bình_thường	Không	Không
E_5	Cao	Không	Không
E_6	Rất_cao	Không	Có
E_7	Cao	Không	Không
E_8	Rất_cao	Có	Có

Định nghĩa 3 [5]. Giả sử O là tập các đối tượng, Ω là tập các thuộc tính, $P, Q \subseteq \Omega$. Độ đo thô, đo mức độ phụ thuộc của tập các thuộc tính Q vào tập các thuộc tính P (được ký hiệu là $\mu_p(Q)$) được xác định như sau:

$$\mu_P(Q) = \frac{\text{card}(\{o \in O \,|\, [o]_P \subseteq [o]_Q\})}{\text{card}(O)}. \tag{4}$$

Khi đó:
- Nếu $\mu_p(Q) = 1$ thì Q phụ thuộc hoàn toàn vào P.
- Nếu $0 < \mu_p(Q) < 1$ thì Q phụ thuộc một phần vào P.
- Nếu $\mu_p(Q) = 0$ thì Q độc lập với P.

Định nghĩa 4 [3]. Giả sử O là tập các đối tượng, Ω là tập các thuộc tính, $P, Q \subseteq \Omega$. Khi đó độ đo R (được ký hiệu bởi $\tilde{\mu}_p(Q)$), đo mức độ phụ thuộc của tập các thuộc tính Q vào tập các thuộc tính P được xác định như sau:

$$\tilde{\mu}_P(Q) = \frac{1}{\text{card}(O)}\left[\sum_{[o]_P} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}([o]_P)}\right]. \tag{5}$$

Tương ứng với dữ liệu trong bảng 1, mức độ phụ thuộc của thuộc tính bị_cúm vào thuộc tính đau_đầu được xác định bởi (5) có giá trị 9/16 trong khi đó độ đo thô tương ứng được xác định bởi (4) có giá trị 0.

Sau đây, chúng ta xây dựng một độ đo mới, độ đo R_N, với ý nghĩa như là một độ đo tin tưởng có giá trị nhỏ hơn độ đo "khả năng" R (trong biểu thức tính trị của R có sử dụng việc lấy giá trị cực đại). Trong biểu thức tính trị của độ đo R_N dưới đây, việc tính trị được thực hiện có dạng "lấy trung bình" theo bình phương.

Định nghĩa 5. Giả sử O là một tập các đối tượng, Ω là tập tất cả các thuộc tính, $P, Q \subseteq \Omega$. Khi đó độ đo R_N (được ký hiệu là ${\mu^N}_P$) đo mức độ phụ thuộc của một tập các thuộc tính Q vào một tập các thuộc tính P được xác định như sau:

$${\mu^N}_P(Q) = \frac{1}{\text{card}(O)}\left[\sum_{[o]_P \subseteq [o]_Q} \text{card}([o]_P) + \sum_{[o]_P \not\subset [o]_Q} \sum_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}^2([o]_P)}\right]. \tag{6}$$

Với dữ liệu trong bảng 1, mức độ phụ thuộc của thuộc tính bị_cúm vào thuộc tính đau_đầu được xác định bởi (6) là 5/32.

4. MỘT SỐ TÍNH CHẤT CỦA ĐỘ ĐO R_N

Mệnh đề 1. *Cho Ω là tập tất cả các thuộc tính. $\forall P, Q \subseteq \Omega$ ta có các đánh giá sau:*

$$\mu_P(Q) \leq {\mu^N}_P(Q) \leq \tilde{\mu}_P(Q).$$

Chứng minh. Trước hết ta viết lại biểu diễn của độ đo thô và độ đo R như sau:

$$\mu_P(Q) = \frac{\text{card}(\{o \in O \mid [o]_P \subseteq [o]_Q\})}{\text{card}(O)} = \frac{1}{\text{card}(O)}\left[\sum_{[o]_P \subseteq [o]_Q} \text{card}([o]_P)\right]. \tag{a}$$

Đặt
$$A = \frac{1}{\text{card}(O)}\left[\sum_{[o]_P \subseteq [o]_Q} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}([o]_P)}\right].$$

Xét $[o]_P \subseteq [o]_Q$, ta cần chỉ ra rằng

$$\max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}([o]_P)} = \text{card}([o]_P). \tag{b}$$

Do $[o]_P \subseteq [o]_Q$ nên tồn tại duy nhất $[o]_Q$ đó để $[o]_Q \cap [o]_P \neq \emptyset$ và giá trị max lấy theo các $[o]_Q$ đạt chính ngay $[o]_Q$ này. Hơn nữa, ta có $\text{card}([o]_Q \cap [o]_P) = \text{card}([o]_P)$ và đẳng thức (b) được kiểm chứng. Vậy

$$\tilde{\mu}_P(Q) = \mu_P(Q) + \frac{1}{\text{card}(O)}\left[\sum_{[o]_P \not\subset [o]_Q} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}([o]_P)}\right]. \tag{c}$$

- Theo (a) và Định nghĩa 5, ta nhận được $\mu_P(Q) \leq {\mu^N}_P(Q)$.
- Ta cần chứng minh về thứ hai ${\mu^N}_P(Q) \leq \tilde{\mu}_P(Q)$.

Theo Định nghĩa 6 và (c), ta chỉ cần chứng minh bất đẳng thức sau:

$$\sum_{[o]_P \not\subset [o]_Q} \sum_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}^2([o]_P)} \leq \sum_{[o]_P \not\subset [o]_Q} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}([o]_P)}. \tag{d}$$

Chúng ta xem xét với từng lớp $[o]_P$, trong trường hợp không tồn tại một lớp $[o]_Q$ nào chứa trọn nó. Khi đó đặt

$$B = \sum_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}^2([o]_P)} = \frac{1}{\text{card}^2([o]_P)} \sum_{[o]_Q} \text{card}^2([o]_Q \cap [o]_P).$$

Vì số lượng các thành phần có giá trị dương tham gia tổng tính B không vượt quá số lượng phần tử của $[o]_P$ (tức là $\text{card}\{[o]_Q : [o]_Q \cap [o]_P \neq \emptyset\} \leq \text{card}([o]_P)$) nên số hạng $\neq 0$ tham gia lấy tổng không vượt quá lực lượng của $[o]_P$. Với mỗi số hạng đó, ta có đánh giá:

$$\text{card}^2([o]_Q \cap [o]_P) \leq \max_{[o]_Q}(\text{card}^2([o]_Q \cap [o]_P))$$

và khi đó

$$\begin{aligned} B &\leq \frac{1}{\text{card}^2([o]_P)} \text{card}([o]_P) \max_{[o]_Q} (\text{card}^2([o]_Q \cap [o]_P)) \\ &= \frac{1}{\text{card}^2([o]_P)} \max_{[o]_Q} (\text{card}^2([o]_Q \cap [o]_P)). \end{aligned}$$

Như vậy, từng thành phần tương ứng 1–1 trong hai vế của (d) đều thỏa mãn dấu bất đẳng thức, và như vậy (d) được chứng minh và $\mu^N{}_P(Q) \leq \tilde{\mu}_P(Q)$. □

Cho Ω là tập tất cả các thuộc tính và hai tập con $P, Q \subseteq \Omega$. Khi xét độ phụ thuộc của tập thuộc tính Q vào tập thuộc tính P, thì P được gọi là tập thuộc tính điều kiện và Q là tập thuộc tính quyết định.

Đối với các luật có dạng "if P then Q", độ tin cậy của chúng phụ thuộc vào sự biến thiên của các tham số P và Q. Sau đây đối với các độ đo sự phụ thuộc thuộc tính, chúng ta khảo sát độ tin cậy của luật này theo hướng cố định tham số quyết định Q và cho biến thiên tham số điều kiện P.

Mệnh đề 2. *Cho Ω là tập tất cả các thuộc tính. $\forall P, Q \subseteq \Omega$ ta luôn có $\tilde{\mu}_P(Q) \leq 1$.*

Chứng minh. $\forall P, Q \subseteq \Omega$, $\forall o \in O$ ta có $([o]_P \cap [o]_Q) \subseteq [o]_P$, $\forall [o]_Q$

$$\begin{aligned} &\Leftrightarrow \quad \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}([o]_P)} \leq \frac{\text{card}^2([o]_P)}{\text{card}([o]_P)} = \text{card}([o]_P) \\ &\Leftrightarrow \quad \tilde{\mu}_P(Q) = \frac{1}{\text{card}(O)} \sum_{[o]_P} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}([o]_P)} \leq \frac{1}{\text{card}(O)} \sum_{[o]_P} \text{card}([o]_P) = 1. \end{aligned}$$ □

Mệnh đề 3. *O là tập các đối tượng, với mọi cặp tập các thuộc tính P, Q ta có khẳng định sau: $\forall o \in O, [o]_P \subseteq [o]_Q$ khi và chỉ khi $\mu_P(Q) = \mu^N{}_P(Q) = \tilde{\mu}_P(Q) = 1$.*

Chứng minh. Đối với độ đo thô của Pawlak, tính đúng đắn của mệnh đề trên là hiển nhiên.

Từ các Mệnh đề 1 và 2, ta có: $\mu_P(Q) \leq \mu^N{}_P(Q) \leq \tilde{\mu}_P(Q) \leq 1$

$$\begin{aligned} &\Rightarrow \quad \forall o \in O, \ [o]_P \subseteq [o]_Q \ \Leftrightarrow \ 1 = \mu_P(Q) \leq \mu^N{}_P(Q) \leq \tilde{\mu}_P(Q) \leq 1 \\ &\Rightarrow \quad \forall o \in O, \ [o]_P \subseteq [o]_Q \ \Leftrightarrow \ 1 = \mu_P(Q) = \mu^N{}_P(Q) = \tilde{\mu}_P(Q) = 1 \end{aligned}$$ □

Hệ quả 1. *Cho Ω là tập tất cả các thuộc tính. Khi đó $\forall Q \subseteq \Omega$*

$$\mu_\Omega(Q) = \mu^N{}_\Omega(Q) = \tilde{\mu}_\Omega(Q) = 1.$$

Định nghĩa 6. Đối với độ đo R_N, $\forall k$ là số thực $0 \leq k \leq 1$, ký hiệu $P \xrightarrow{k}_{R_N} Q$ được định nghĩa là Q phụ thuộc độ k vào P nếu như $k = \mu^N{}_P(Q)$.

- Nếu $k = 1$, nói rằng *Q phụ thuộc hoàn toàn vào P* (ký hiệu $P \longrightarrow_{R_N} (Q)$.
- Nếu $0 < k < 1$ nói rằng Q phụ thuộc độ k vào P (phụ thuộc một phần).
- Nếu $k = 0$ nói rằng Q độc lập với P.

Bổ đề 1. *$\forall$ số nguyên a_i, b_i với a_i không âm và b_i dương ($i = 1, 2, \ldots, n$), ta có*

$$\frac{\left(\sum_{i=1}^n a_i\right)^2}{\sum_{i=1}^n b_i} \leq \sum_{i=1}^n \frac{a_i^2}{b_i} \quad \text{và} \quad \frac{\left(\sum_{i=1}^n a_i\right)^2}{\left(\sum_{i=1}^n b_i\right)^2} \leq \sum_{i=1}^n \frac{a_i^2}{b_i^2}.$$

Chứng minh. Bổ đề là hệ quả của bất đẳng thức Buniacovski. □

Định lý 1. *Độ đo thô của Pawlak, độ đo R, độ đo R_N thỏa mãn tiên đề đơn điệu.*

Chứng minh. Xét hai tập thuộc tính P và P' trong đó $P \subseteq P'$. Gọi m là một độ đo trong ba độ đo nói trên, ta cần chứng minh $m(P') \geq m(P)$.

Chú ý mở đầu:

Giả sử rằng tập đối tượng O được phân hoạch theo tập thuộc tính P thành q lớp tương đương. Do $P \subseteq P'$ nên mỗi lớp tương đương thứ i theo tập thuộc tính P sẽ bao gồm n_i $(i = 1, 2, \ldots, q)$ lớp tương đương theo tập thuộc tính P'.

Ký hiệu đối tượng đại diện cho lớp tương đương thứ j $(j = 1, 2, \ldots, n_i)$ theo tập thuộc tính P' nằm gọn trong lớp tương đương thứ i theo tập thuộc tính P là ${o_i}^j$ $(j = 1, 2, \ldots, n_i)$. Với mỗi lớp tương đương thứ i theo tập thuộc tính P, ký hiệu đối tượng đại diện là ${o_i}^*$. Ta có thể chọn các phần tử ${o_i}^*$ từ một trong các phần tử ${o_i}^j$ trong một số trường hợp nào đó mà không làm giảm tính tổng quát của các chứng minh.

Xét một lớp tương đương thứ i (tức là $[{o_i}^*]_P$) theo tập thuộc tính P ta có:

(i1) $[o_i]_P = \sum_{j=1}^{n_i} [{o_i}^j]_{P'}$.

(i2) $\text{card}([o_i]_P) = \sum_{j=1}^{n_i} \text{card}([{o_i}^j]_{P'})$.

(i3) Với lớp tương đương $[o]_Q$ bất kỳ theo tập thuộc tính Q, luôn có:

$$\text{card}([o]_Q \cap [o_i]_P) = \sum_{j=1}^{n_i} \text{card}([o]_Q \cap [{o_i}^j]_{P'}).$$

- m là độ đo thô:

Xét hai tập hợp $O_1 = \{o \in O : [o]_P \subseteq [o]_Q\}$ và $O_2 = \{o \in O : [o]_{P'} \subseteq [o]_Q\}$.

Với bất kỳ $o \in O_1$, xét lớp tương đương $[o]_P$. Theo trên có $o = o_i$ nào đó và $[o_i]_{P'} \subseteq [o_i]_P \subseteq [o_i]_Q$, như vậy $o \in O_2$. Do o bất kỳ nên có $O_1 \subseteq O_2$.

Từ đó $\text{card}(O_1) \leq \text{card}(O_2)$ hay $m(P) \leq m(P')$.

- m là độ đo R:

Theo chú ý mở đầu, chúng ta có các đẳng thức sau đây:

$$\text{card}(O) \times \tilde{\mu}_P(Q) = \sum_{[o]_P} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}([o]_P)} = \sum_{i=1}^{q} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [{o_i}^*]_P)}{\text{card}([{o_i}^*]_P)}$$

và

$$\text{card}(O) \times \tilde{\mu}_{P'}(Q) = \sum_{[o]_{P'}} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_{P'})}{\text{card}([o]_{P'})} = \sum_{i=1}^{q} \sum_{j=1}^{n_i} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [{o_i}^j]_{P'})}{\text{card}([{o_i}^j]_{P'})}$$

Do $\text{card}(O)$ cố định nên để chứng minh $\tilde{\mu}_P(Q) \leq \tilde{\mu}_{P'}(Q)$ ta chỉ cần chứng minh

$$\sum_{i=1}^{q} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [{o_i}^*]_P)}{\text{card}([{o_i}^*]_P)} \leq \sum_{i=1}^{q} \sum_{j=1}^{n_i} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [{o_i}^j]_{P'})}{\text{card}([{o_i}^j]_{P'})} \qquad \text{(e)}$$

Hai vế của (e) cùng có q số hạng nên để kiểm chứng bất đẳng thức này, chúng ta chỉ cần kiểm chứng từng cặp số hạng tươngứng trong q số hạng này. Tức là ta phải chứng minh được với $i = 1, 2, \ldots, q$:

$$\max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [{o_i}^*]_P)}{\text{card}([{o_i}^*]_P)} \leq \sum_{j=1}^{n_i} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o_i^j]_{P'})}{\text{card}([o_i^j]_{P'})} \qquad \text{(g)}$$

Cũng theo chú ý mở đầu, có thể chọn ${o_i}^*$ làm phần tử đại diện cho lớp tương đương thứ i theo tập thuộc tính P với một số tính chất đặc biệt nào đó. Không làm giảm tổng quát, chọn o_i^* là chính

là phần tử làm cực đại $\dfrac{\text{card}^2([o]_Q \cap [o_i{}^*]_P)}{\text{card}([o_i{}^*]_P)}$ (thuộc lớp tương đương theo tập thuộc tính Q làm cực đại). Do $o_i{}^*$ đã được chọn trên đây thuộc vào $[o]_P$ mà $[o]_P$ phân hoạch thành các $[o_i{}^j]_{P'}$ nên $o_i{}^*$ thuộc vào lớp tương dương thứ j_0 nào đó: lớp tương đương $[o_i^{j_0}]_{P'}$. Như vậy không làm giảm tổng quá ta chọn phần tử đại diện $o_i^* = o_i^{j_0}$ có lớp tương đương theo $Q([o_i^{j_0}])_Q)$ làm cực đại về trái của (g).

Như vậy, về trái của (g) có giá trị chính là

$$\frac{\text{card}^2([o_i^{j_0}]_Q \cap [o_i^{j_0}]_P)}{\text{card}([o_i^{j_0}]_P)} \tag{h}$$

Đối với về phải của (g), với $j = 1, 2, \ldots, n_i$, chúng ta luôn có:

$$\max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o_i^j]_{P'})}{\text{card}([o_i^j]_{P'})} \geq \frac{\text{card}^2([o_i^{j_0}]_Q \cap [o_i^j]_{P'})}{\text{card}([o_i^j]_{P'})}$$

và như vậy

$$\sum_{j=1}^{n_i} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o_i^j]_{P'})}{\text{card}([o_i^j]_{P'})} \geq \sum_{j=1}^{n_i} \frac{\text{card}^2([o_i^{j_0}]_Q \cap [o_i^j]_{P'})}{\text{card}([o_i^j]_{P'})} = B.$$

Theo bất đẳng thức thứ nhất của Bổ đề 1, ta nhận được:

$$B \geq \frac{\left(\sum_{j=1}^{n_i} \text{card}([o_i^{j_0}]_Q \cap [o_i^j]_{P'})\right)^2}{\sum_{j=1}^{n_i} \text{card}([o_i{}^j]_{P'})} = \frac{\text{card}^2([o_i^{j_0}]_Q \cap [o_i^{j_0}]_P)}{\text{card}([o_i^j]_P)}$$

(theo các hệ thức (i2) và (i3) trong chú ý mở đầu và chọn ngay $o_i^{j_0}$ làm phần tử đại diện $o_i{}^*$ trong lớp tương đương theo P).

Vậy

$$\sum_{j=1}^{n_i} \max_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o_i{}^j]_{P'})}{\text{card}([o_i{}^j]_{P'})} \geq \frac{\text{card}^2([o_i^{j_0}]_Q \cap [o_i^{j_0}]_P)}{\text{card}([o_i^{j_0}]_P)}.$$

Như vậy, (g) được kiểm tra đúng với mọi số hạng thứ i $n(i = 1, 2, \ldots, q)$ có nghĩa là $m(P) \leq m(P')$ hay cũng vậy $R(P) \leq R(P')$.

- m là độ đo R_N:

Tương tự như trên, ta xét:

$$\text{card}(O) \times \mu^N{}_P(Q) = \sum_{[o]_P \subseteq [o]_Q} \text{card}([o]_P) + \sum_{[o]_P \not\subset [o]_Q} \sum_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}^2([o]_P)}$$

$$= A + \sum_{[o]_P \not\subset [o]_Q} \sum_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}^2([o]_P)}$$

với $A = \sum_{[o]_P \subseteq [o]_Q} \text{card}([o]_P)$

và

$$\text{card}(O) \times \mu^N_{P'}(Q) = \sum_{[o]_P \subseteq [o]_Q} \text{card}([o]_{P'}) + \sum_{[o]_P \not\subset [o]_Q} \sum_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_{P'})}{\text{card}^2([o]_{P'})}.$$

Do $\text{card}(O)$ cố định nên để chứng minh $\mu^N_P(Q) \leq \mu^N_{P'}(Q)$ ta chỉ cần chứng minh quan hệ nói trên đối với hai vế phải của hai biểu diễn trên.

Ta nhận được đánh giá sau:

$$\forall [o]_P \text{ luôn có } \sum_{[o]_Q} \frac{\text{card}^2([o]_Q \cap [o]_P)}{\text{card}^2([o]_P)} \leq \text{card}([o]_P) \tag{i}$$

do

$$\operatorname{card}([o]_P) = \sum_{[o]_Q} \operatorname{card}([o]_Q \cap [o]_P)$$

và

$$\frac{\operatorname{card}^2([o]_Q \cap [o]_P)}{\operatorname{card}^2([o]_P)} = \operatorname{card}([o]_Q \cap [o]_P)\frac{\operatorname{card}([o]_Q \cap [o]_P)}{\operatorname{card}^2([o]_P)} \leq \operatorname{card}([o]_Q \cap [o]_P).$$

Phân loại các lớp tương đương được phân hoạch bởi tập thuộc tính P' thành ba loại như sau:

+ $[o]_{P'} \subseteq [o]_P \subseteq [o]_Q$ là các lớp tương đương được chia từ các lớp tương đương theo phân hoạch P đều tương ứng có các lớp tương đương tham gia tổng A theo phân hoạch P' thuộc loại này và sẽ cho tổng lực lượng như nhau. Gọi tập gồm các lớp tương đương $[o]_{P'}$ thuộc loại này là tập I.

+ $[o]_{P'} \subseteq [o]_P$ song $[o]_{P'} \not\subseteq [o]_Q$. Hạng thức khi tính giá trị độ đo tương ứng với lớp tương đương này theo P' sẽ là card $([o]_{P'})$. Gọi tập hợp gồm các lớp tương đương $[o]_{P'}$, thuộc loại này là II.

+ $[o]_{P'} \not\subset [o]_Q$: Gọi tập hợp gồm các lớp tương đương $[o]_{P'}$ thuộc loại này là III.

Liên hệ với chú ý mở đầu và không làm giảm tổng quát ta giả thiết rằng các lớp tương đương theo tập P tương ứng với các lớp tương đương theo P' trong tập I là các lớp $[o_i^*]_P$ đầu tiên ($i = 1, 2, \ldots, k$; với $0 \leq k \leq q$). Để ý rằng, khi $k = 0$ thì không có bất kỳ một lớp tương đương theo tập P nằm trọn trong một lớp tương đương theo tập Q; còn khi $k = q$ thì mọi lớp tương đương theo tập P đều nằm trọn trong một lớp tương đương nào đó theo tập Q.

Có thể viết lại card $(O) \times \mu_P^N(Q)$ như sau:

$$\begin{aligned}\operatorname{card}(O) \times \mu_P^N(Q) &= \sum_{i=1}^{k} \operatorname{card}([o_i^*]_P) + \sum_{i=k+1}^{q} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o_i^*]_P)}{\operatorname{card}^2([o_i^*]_P)} \\ &= A + \sum_{i=k+1}^{q} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o_i^*]_P)}{\operatorname{card}^2([o_i^*]_P)} \qquad \text{(j)}\end{aligned}$$

Chúng ta xét tổng sau liên quan đến tập thuộc tính P':

$$\begin{aligned}\operatorname{card}(O) \times \mu_{P'}^N(Q) &= \sum_{[o]_{P'} \subseteq [o]_Q} \operatorname{card}([o]_{P'}) + \sum_{[o]_{P'} \not\subset [o]_Q} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o]_{P'})}{\operatorname{card}^2([o]_{P'})} \\ &= \sum_{[o]_{P'} \in \mathrm{I}} \operatorname{card}([o]_{P'}) + \sum_{[o]_{P'} \in \mathrm{II}} \operatorname{card}([o]_{P'}) + \sum_{[o]_{P'} \in \mathrm{III}} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o]_{P'})}{\operatorname{card}^2([o]_{P'})} \\ &= \sum_{i=1}^{k} \operatorname{card}([o_i^*]_{P'}) + \sum_{[o]_{P'} \in \mathrm{II}} \operatorname{card}([o]_{P'}) + \sum_{[o]_{P'} \in \mathrm{III}} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o]_{P'})}{\operatorname{card}^2([o]_{P'})} \\ &= A + \sum_{[o]_{P'} \in \mathrm{II}} \operatorname{card}([o]_{P'}) + \sum_{[o]_{P'} \in \mathrm{III}} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o]_{P'})}{\operatorname{card}^2([o]_{P'})} \\ &\text{(theo (i))} \geq A + \sum_{[o]_{P'} \in \mathrm{II} \cup \mathrm{III}} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o]_{P'})}{\operatorname{card}^2([o]_{P'})} = A + C. \qquad \text{(k)}\end{aligned}$$

với $C = \sum_{[o]_{P'} \in \mathrm{II} \cup \mathrm{III}} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o]_{P'})}{\operatorname{card}^2([o]_{P'})}$.

Sau khi nhóm lại các lớp tương đương theo tập thuộc tính P' thành các lớp tương đương theo tập thuộc tính P, ta có:

$$C = \sum_{i=k+1}^{q} \sum_{j=1}^{n_i} \sum_{[o]_Q} \frac{\operatorname{card}^2([o]_Q \cap [o_i^j]_{P'})}{\operatorname{card}^2([o_i^j]_{P'})} = \sum_{i=k+1}^{q} \sum_{[o]_Q} \sum_{j=1}^{n_i} \frac{\operatorname{card}^2([o]_Q \cap [o_i^j]_{P'})}{\operatorname{card}^2([o_i^j]_{P'})}.$$

Từ bất đẳng thức thứ hai trong Bổ đề 1 và chú ý mở đầu ((i2), (i3), ta có:

$$\sum_{j=1}^{n_i} \frac{\mathrm{card}^2([o]_Q \cap [o_i^j]_{P'})}{\mathrm{card}^2([o_i^j]_{P'})} \geq \frac{\left(\sum\limits_{j=1}^{n_i} \mathrm{card}^2([o]_Q \cap [o_i^j]_{P'})\right)^2}{\left(\mathrm{card}^2([o_i^j]_{P'})\right)^2} = \frac{\mathrm{card}^2([o]_Q \cap [o_i^*]_P)}{\mathrm{card}^2([o_i^*]_P)}$$

và nhận được

$$C \geq \sum_{i=k+1}^{q} \sum_{[o])Q} \frac{\mathrm{card}^2([o]_Q \cap [o_i^*]_P)}{\mathrm{card}^2([o_i^*]_P)}. \tag{l}$$

Từ (j), (k) và (l) ta có $R_N(P) \leq R_N(P')$. □

Từ Hệ quả 1 và Định lý 1 ta thấy rằng: nếu coi tập tất cả các thuộc tính Ω là tập tham chiếu thì độ đo thô của Pawlak, độ đo R, độ đo R_N là các độ đo tin tưởng.

Mệnh đề 4. *$\forall P, Q \subseteq \Omega$, $(P \cap Q) = \emptyset$, ký hiệu $\overline{P}$ là phần bù của P trong Ω, khi đó:*

$$\mu_{\overline{P}}(Q) = \mu^N{}_{\overline{P}}(Q) = \tilde{\mu}_{\overline{P}}(Q) = 1.$$

Chứng minh. Suy từ Mệnh đề 3. □

Tương tự các kết quả về sự phụ thuộc thô trong [2], chúng ta có các Mệnh đề 5 và Mệnh đề 6 như dưới đây.

Mệnh đề 5. *Đối với độ đo R_N ta có các tính chất sau:*

(1) *Nếu $B \supseteq C$ thì $B \longrightarrow_{R_N} C$,*

(2) *Nếu $B \longrightarrow_{R_N} C$ thì $\forall D \subseteq \Omega$ đều có $BD \longrightarrow_{R_N} CD$,*

(3) *Nếu $B \longrightarrow_{R_N} C$ và nếu $C \longrightarrow_{R_N} D$ thì $B \longrightarrow_{R_N} D$.*

Chứng minh.

(1): Do $B \supseteq C$ ta có $[o]_B \subseteq [o]_C \Rightarrow B \longrightarrow_{R_N} C$ (Mệnh đề 3).

(2): Từ $B \longrightarrow_{R_N} C \Rightarrow [o]_B \subseteq [o]_C$ (Mệnh đề 3) $\Rightarrow [o]_{BD} \subseteq [o]_{CD} \Rightarrow BD \longrightarrow_{R_N} CD$.

(3): Do $B \longrightarrow_{R_N} C$ và $C \longrightarrow_{R_N} D \Rightarrow [o]_B \subseteq [o]_C$ và $[o]_C \subseteq [o]_D \Rightarrow [o]_B \subseteq [o]_D \Rightarrow B \longrightarrow_{R_N} D$.

□

Mệnh đề 6. *Cho Ω là tập tất cả các thuộc tính. Đối với độ đo phụ thuộc thuộc tính R_N thì các khẳng định sau đây không đúng:*

(1) *Nếu $B \xrightarrow{k}_{R_N} C$ và $\forall D \subseteq \Omega$ thì $BD \xrightarrow{k}_{R_N} CD$,*

(2) *Nếu ($B \xrightarrow{k}_{R_N} C$ và $C \longrightarrow_{R_N} D$) hoặc ($B \longrightarrow_{R_N} C$ và $C \xrightarrow{k}_{R_N} D$) thì $B \xrightarrow{k}_{R_N} D$.*

Chứng minh. Để chứng minh mệnh đề trên, chúng ta sử dụng phương pháp phản chứng thông qua việc chỉ ra các phản ví dụ. Xét tập các đối tượng nào đó (mỗi đối tượng có thông tin thể hiện một hàng) với các thuộc tính $A < B, C$ như sau:

A	B	C
1	1	1
1	2	1
1	2	2

(1) $\mu_{N\,A}(C) = (1 + 4/9 + 1/9)/4 = 7/18$ hay $A \xrightarrow{7/18}_{R_N} C$,

$\mu^N{}_{(A \cup B)}(C \cup B) = (1 + 1/4 + 1/4 + 1)/4 = 5/8$ hay $AB \xrightarrow{5/8}_{R_N} CB \Rightarrow$ (1) được chứng minh.

(2) Đối với trường hợp thứ nhất, ta có

$\mu_{NC}(B) = (1/4 + 1/4 + 1/4 + 1/4)/4 = 1/4$ hay $C \xrightarrow{1/4}_{R_N} B$,
$[o]_B \subseteq [o]_A \Rightarrow B \longrightarrow_{R_N} A$,
$\mu_{NC}(A) = (1 + 1 + 1/4 + 1/4)/4 = 5/8$ hay $C \xrightarrow{5/8}_{R_N} A$.
Đối với trường hợp thứ hai, ta có:
$[o]_B \subseteq [o]_A \Rightarrow B \longrightarrow_{R_N} A$,
$\mu_{NA}(C) = (1 + 4/9 + 1/9)/4 = 7/18$ hay $A \xrightarrow{7/18}_{R_N} C$,
$\mu_{NB}(C) = (1 + 1 + 1/4 + 1/4)/4 = 5/8$ hay $B \xrightarrow{5/8}_{R_N} C$. □

5. BÀN LUẬN

Theo Dubois và Prade [1], một cặp các độ đo tin tưởng đối ngẫu nhau thường được cùng xem xét như là cặp hai độ đo ngưỡng: độ đo cần thiết N và độ đo khả năng Π. Độ đo cần thiết N được xem như độ tin cậy tối thiểu có được còn độ đo khả năng Π được xem như độ tin cậy tối đa. Nằm giữa hai độ đo nói trên là một lớp độ đo tin cậy mà trong đó có độ đo xác suất. Chúng ta có thể coi hai độ đo R và độ đo thô là hai độ đo ngưỡng theo một ngữ cảnh đặc biệt nào đó và R_N như một độ đo tin cậy nằm giữa chúng (Mệnh đề 1) trong cùng ngữ cảnh. Tuy nhiên hai độ đo được coi là ngưỡng như giới thiệu ở đây thực sự không có mối quan hệ mật thiết như hai độ đo Π và N.

TÀI LIỆU THAM KHẢO

[1] Dubois Didier, Prade Henri, Possibility theory: An approach to computerized processing of uncertainly, *CNSR, Languages and Computer System (LSI)*, University of Toulouse III, 1986. (Bản dịch tiếng Anh do University of Cambridge, 1988).
[2] Hà Quang Thụy, Tập thô trong bảng quyết định, *Tạp chí Khoa học Đại học Quốc gia Hà Nội* **12** (4) (1996) 9-14.
[3] Ho Tu Bao and Nguyen Trong Dung, A rough sets based measure for workshop on rough sets, *Fuzzy Sets and Machine Discovery (RSFD '96)*, 1996.
[4] Le Tien Vuong and Ho Thuan, A relation database extended by applications of fuzzy set theory and linguistic variables, *Computers and Artificial Intelligence, Bratislava* **9** (2) (1989) 153-168.
[5] Pawlak Z., Rough set and decision tables, *ICS PAS Report, Warsawa, Poland* **540** (3) (1984).
[6] Theresa Beaubouef, Frederik E., and Gurdial Aroza, Informationtheoretic measures of uncertainty for rough sets and tough relational databases, *Journal of Information Science* **409** (1998) 185-195.

Nhận bài ngày 10 - 9 - 1999
Nhận lại sau khi sửa ngày 20 - 4 - 2000

Đỗ Tấn Phong - Công ty Điện thoại di động VMS.
Hồ Thuần - Viện Công nghệ thông tin.
Hà Quang Thụy - Trường Đại học Khoa học tự nhiên.