Tạp chí Tin học và Điều khiển học, T. 20, S. 1 (2004), 335–346

MỘT SỐ TÍNH TOÁN CHO TỔNG KẾT DỮ LIỆU TRÊN CƠ SỞ DỮ LIỆU QUAN HỆ MỜ

TRẦN THIÊN THÀNH

Trường Đại học Sư phạm Quy Nhơn

Abstract. In this paper, we present some computations of data summary rules on Fuzzy relational databases based on a pattern matching process of D. Dubois and H. Prade. We take advantage of some properties of these rules to design algorithms involving heuristics. Some algorithms (with heuristics) which allow us to determine whether, for the value of truth of data summary rules is greater than a given threshold. The correctness and time complexity of these algorithms are given.

Tóm tắt. Bài báo trình bày một số kết quả tính toán độ tin cậy cho các luật tổng kết từ dữ liệu trên mô hình cơ sở dữ liệu quan hệ mờ. Các tính toán được xây dựng dựa trên cách tiếp cận của D. Dubois và H. Prade. Bài báo cũng đưa ra một số tính chất của những luật này nhằm mục đích thiết kế các thuật toán heuristics để kiểm tra độ tin cậy của các luật với ngưỡng cho trước. Tính đúng đắn và độ phức tạp của các thuật toán cũng được trình bày.

1. ĐẶT VẤN ĐỀ

Trong [11], chúng tôi đã xây dựng cách đánh giá độ tin cậy cho các luật có dạng "$Q\ P_1\ r\ are\ P_2$" và "$Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r$", trong đó Q, Q_1, Q_2 là các lượng từ mờ, r là một quan hệ với dữ liệu mờ, P_1, P_2 là các tân từ mờ, θ là toán tử so sánh mờ. "$Q\ P_1\ r\ are\ P_2$" có nghĩa là định lượng các bộ trong quan hệ r thoả tân từ P_1 cũng thoả tân từ P_2 tương thích với lượng từ Q. "$Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r$" có ý nghĩa là định lượng các bộ trong quan hệ r thoả tân từ P_1 có quan hệ θ với các bộ trong quan hệ r thoả tân từ P_2 ở mức độ của lượng từ Q_2 là tương thích với lượng từ Q_1. Cùng với cách đánh giá độ tin cậy, dựa vào thứ tự phân cấp của các tập mờ, chúng tôi đưa ra thuật toán xây dựng tập các luật tổng kết dữ liệu theo mẫu cho trước trên các dữ liệu có sẵn. Các thuật toán tổng kết dữ liệu phụ thuộc nhiều vào việc kiểm tra độ tin cậy của các luật với một ngưỡng cho trước. Các thuật toán này thường có độ phức tạp khá lớn do công thức tính toán phức tạp và lượng dữ liệu cần tính toán lớn. Trong bài báo này, với mục đích cải tiến các thuật toán kiểm tra độ tin cậy các luật tổng kết từ dữ liệu, chúng tôi đưa ra các tính chất của tính toán độ tin cậy dựa vào đặc trưng của các luật. Từ đó xây dựng các thuật toán heuristics kiểm tra độ tin cậy của các luật với ngưỡng cho trước hiệu quả hơn. Trong các luật tổng kết dữ liệu, các nghiên cứu trước đây thường dùng những lượng từ đơn điệu, chúng tôi đưa ra các kết quả đánh giá cho lượng từ unimodal thông qua các lượng từ đơn điệu.

Mục 2 sẽ trình bầy những kiến thức cơ sở gồm: Đánh giá các tân từ mờ; Biểu diễn các lượng từ ngôn ngữ bằng tập mờ; Lực lượng tập mờ. Mục 3 trình bầy các tính chất của từng dạng luật và thuật toán tính, kiểm tra độ tin cậy của các luật với ngưỡng cho trước. Cuối cùng là kết luận và một số hướng nghiên cứu tiếp theo.

* Công trình này được hoàn thành với sự hỗ trợ kinh phí của đề tài cấp Nhà nước mã số 230203.

2. CÁC KIẾN THỨC CƠ SỞ

2.1. Tính toán trên các tân từ mờ

Cho X là một biến nhận giá trị trên miền D kết hợp với phân bố khả năng π_X, F là một tập mờ trên D. Độ tương thích của X với tập mờ F được đánh giá trên hai độ đo khả năng (Π) và cần thiết (N) được xác định bởi:

$$\Pi_X(F) = \sup_{u\in D} \min\Big(\mu_F(u), \pi_X(u)\Big) \tag{2.1}$$

$$N_X(F) = \inf_{u\in D} \max\Big(\mu_F(u), 1-\pi_X(u)\Big). \tag{2.2}$$

với θ là một phép so sánh mờ được xác định bởi hàm thuộc μ_θ, mệnh đề "$X\ \theta\ F$" được xem tương đương với mệnh đề "$X\ is\ F\circ\theta$", với $F\circ\theta$ là phép hợp thành của một giá trị mờ F với một toán tử so sánh mờ θ được xác định bởi: $\forall d\in D, \mu_{F\circ\theta}(d) = \sup_{d'\in D} \min\Big(\mu_\theta(d,d'), \mu_F(d')\Big)$.

2.2. Các lượng từ mờ

Trong [14] Zadeh đề xuất một cách biểu diễn lượng từ ngôn ngữ theo cách tiếp cận của lý thuyết tập mờ. Trong đó mỗi lượng từ được xem như một tập mờ. Zadeh chia các lượng từ ngôn ngữ thành hai loại: *lượng từ tuyệt đối* (absolute quantifiers) và *lượng từ tỷ lệ* (proportional quantifiers). Dựa vào hàm thuộc của lượng từ ta gọi Q là lượng từ đơn điệu tăng (hoặc giảm) nếu hàm thuộc của nó là hàm tăng (hay giảm).

Lượng từ Q gọi là lượng từ *unimodal* nếu tồn tại hai giá trị a, b với $a \le b$ sao cho với mọi $x < a$ thì Q là lượng từ đơn điệu tăng; với $x > b$ thì Q đơn điệu giảm và $\mu_Q(x) = 1$ với mọi $x \in [a, b]$. Với mọi lượng từ unimodal Q bao giờ cũng tìm được hai lượng từ Q_a đơn điệu tăng và Q_d đơn điệu giảm sao cho $Q = Q_a \cap Q_d$. Thật vậy, dễ dàng chỉ ra các lượng từ đơn điệu Q_a và Q_d xác định như sau: $\mu_{Q_a}(x) = \mu_Q(x)$, với $x \le a$; 1, nếu $x > a$, $\mu_{Q_d}(x) = \mu_Q(x)$, với $x \ge b$; 1, nếu $x < b$.

2.3. Lực lượng tập mờ

Cho F là tập mờ trên tập hữu hạn $U = \{u_1, u_2, \cdots, u_n\}$. Đặt $k = |\ker(F)|$, với $\ker(F) = \{u \in U | \mu_F(u) = 1\}$. Lực lượng của tập mờ F, ký hiệu $|F|_f$ (hoặc $|F|$ nếu không gây nhầm lẫn) là một phân bố khả năng $\pi_{|F|}$ trên đoạn $[0, n]$, được xác định như sau:

$\pi(t) = 0$ với $0 \le t < k$,

$\pi(k) = 1$,

với $j > k$ thì $\pi(j)$ là giá trị lớn thứ j trong danh sách các giá trị $\mu(u_1), \mu(u_2), \cdots, \mu(u_n)$.

3. TÍNH TOÁN ĐỘ TIN CẬY CỦA CÁC LUẬT TỔNG KẾT DỮ LIỆU

Trong phần này xây dựng các thuật toán tính và kiểm tra độ tin cậy của các luật với ngưỡng cho trước. Các thuật toán này được xây dựng dựa vào một số tính chất của từng dạng luật. Độ tin cậy được đánh giá trên hai độ đo khả năng và cần thiết.

3.1. Dạng *Q r are P*

Trong [1], độ tin cậy của mệnh đề "$Q\ r\ are\ P$" được đánh giá:

$$\Pi_{Q\ r\ are\ P} = \max_{k\le i\le n} \Big\{\min\Big(\mu_Q(i), \pi_{|r_P|}(i)\Big)\Big\} \tag{3.1}$$

$$N_{Q\ r\ are\ P} = \min_{k\le i\le n} \Big\{\max\Big(\mu_Q(i), 1-\pi_{|r_P|}(i)\Big)\Big\}, \tag{3.2}$$

với n là số bộ của quan hệ r, $k = |\ker(r_P)|$, r_p là quan hệ mờ gồm các bộ của quan hệ r với độ thoả lượng từ P cho từng bộ.

Nếu Q là lượng từ tỷ lệ, trong công thức trên ta thay $\mu_Q(i)$ bởi $\mu_Q(i/n)$.

Trong [1] đã đưa ra công thức tốt cho việc tính toán độ tin cậy của các luật dạng này với các lượng từ đơn điệu. Cụ thể trong trường hợp Q là lượng từ đơn điệu tăng thì

$$\Pi_{Q\ r\ are\ P} = \mathrm{med}\Big(\mu_Q(1), \mu_Q(2), \cdots, \mu_Q(m-1), \mu_P(t_1), \mu_P(t_2), \cdots, \mu_P(t_m)\Big) \tag{3.3}$$

$$N_{Q\ r\ are\ P} = \mu_Q(k) \tag{3.4}$$

với m là số nguyên thoả $m \geq n$, $\mu_Q(m) = 1$, nếu $m > n$ thì $t_{n+1}, \cdots, t_m$ là các bộ ảo bổ sung vào quan hệ r, các bộ này có độ thoả tân từ P là 0 ($\mu_P(t_i) = 0, \forall i, n+1 \leq i \leq m$).

Trong trường hợp Q là lượng từ đơn điệu giảm thì ta cũng có các kết quả tương tự:

$$\Pi_{Q\ r\ are\ P} = \mu_Q(k) \tag{3.5}$$

$$N_{Q\ r\ are\ P} = \mathrm{med}\Big(\mu_Q(1), \cdots, \mu_Q(m-1), 1-\mu_P(t_1), \cdots, 1-\mu_P(t_m)\Big) \tag{3.6}$$

Chúng tôi phát triển các kết quả tính toán cho trường hợp Q là lượng từ unimodal.

Định lý 3.1. *Cho Q là một lượng từ unimodal được phân tích thành giao của hai lượng từ đơn điệu $Q = Q_a \cap Q_d$. Khi đó với mọi quan hệ r và mọi tân từ P trên r, ta có:*

$$\Pi_{Q\ r\ are\ P} = \min\Big(\Pi_{Q_a\ r\ are\ P}, \Pi_{Q_d\ r\ are\ P}\Big) \text{ và } N_{Q\ r\ are\ P} = \min\Big(N_{Q_a\ r\ are\ P}, N_{Q_d\ r\ are\ P}\Big).$$

Chứng minh. Đặt $k = |\ker(r_P)|$.

Trước hết ta chứng minh

$$\Pi_{Q\ r\ are\ P} = \min\Big(\Pi_{Q_a\ r\ are\ P}, \Pi_{Q_d\ r\ are\ P}\Big) \tag{*}$$

Vì $Q \subseteq Q_a$ và $Q \subseteq Q_d$ nên $\Pi_{Q\ r\ are\ P} \leq \min\Big(\Pi_{Q_a\ r\ are\ P}, \Pi_{Q_d\ r\ are\ P}\Big)$.

Ta cần chứng minh $\Pi_{Q\ r\ are\ P} \geq \min\Big(\Pi_{Q_a\ r\ are\ P}, \Pi_{Q_d\ r\ are\ P}\Big)$. Xét các trường hợp sau:

Trường hợp 1: $k \leq a$.

Vì $k \leq a$ nên $\forall k \leq i \leq a$ ta có $\mu_{Q_a}(i) = \mu_Q(i)$, và $\forall n \geq i > a$ thì $\mu_{Q_a}(i) = 1$.

Từ đó suy ra $\Pi_{Q_a\ r\ are\ P} = \max\limits_{k \leq i \leq a}\Big\{\min\Big(\mu_Q(i), \pi_{|r_P|}(i)\Big)\Big\} \leq \Pi_{Q\ r\ are\ P}$.

Do đó, nếu $k \leq a$ thì $\Pi_{Q\ r\ are\ P} \geq \min\Big(\Pi_{Q_a\ r\ are\ P}, \Pi_{Q_d\ r\ are\ P}\Big)$.

Trường hợp 2: $k > a$.

Khi đó $\Pi_{Q_d\ r\ are\ P} = \mu_{Q_d}(k) = \mu_Q(k) = \min\Big(\mu_Q(k), \pi_{|r_P|}(k)\Big) \leq \Pi_{Q\ r\ are\ P}$.

Do đó $\Pi_{Q\ r\ are\ P} \geq \min\Big(\Pi_{Q_a\ r\ are\ P}, \Pi_{Q_d\ r\ are\ P}\Big)$.

Vậy $\Pi_{Q\ r\ are\ P} = \min\Big(\Pi_{Q_a\ r\ are\ P}, \Pi_{Q_d\ r\ are\ P}\Big)$. Công thức (*) được chứng minh.

Tiếp theo, ta chứng minh $N_{Q\ r\ are\ P} = \min\Big(N_{Q_a\ r\ are\ P}, N_{Q_d\ r\ are\ P}\Big)$ (**)

Ta có

$$N_{Q\ r\ are\ P} = \min\Bigg(\min_{k \leq i \leq b}\Big\{\max\Big(\mu_Q(i), 1-\pi_{|r_P|}(i)\Big)\Big\}, \min_{b < i \leq n}\Big\{\max\Big(\mu_Q(i), 1-\pi_{|r_P|}(i)\Big)\Big\}\Bigg).$$

Vì $\forall i > b$ thì $\mu_{Q_a}(i) = 1$, và $\forall i \leq b$ thì $\mu_{Q_d}(i) = 1$ nên ta có

$$N_{Q\ r\ are\ P} = \min\Big(N_{Q_a\ r\ are\ P}, N_{Q_d\ r\ are\ P}\Big).$$

Công thức (**) được chứng minh. ∎

Hệ quả 3.1. *Với giả thiết như định lý trên, ta có*

$$\Pi_{Q\ r\ are\ P} = \begin{cases} \Pi_{Q_a\ r\ are\ P} & \textit{nếu } k < a \\ 1 & \textit{nếu } a \leq k \leq b \\ \mu_Q(k) & \textit{nếu } k > b \end{cases} \tag{1}$$

$$N_{Q\ r\ are\ P} = \begin{cases} \min\Big(\mu_Q(k), N_{Q_d\ r\ are\ P}\Big) & \textit{nếu } k \leq a \\ N_{Q_d\ r\ are\ P} & \textit{nếu } k > a \end{cases} \tag{2}$$

Chứng minh. Đầu tiên ta chứng minh công thức (1).

Nếu $k < a$ thì $\Pi_{Q_d\ r\ are\ P} = \mu_{Q_d}(k) = 1$.

Theo Định lý 3.1, ta có $\Pi_{Q\ r\ are\ P} = \Pi_{Q_a\ r\ are\ P}$.

Nếu $a \leq k \leq b$ thì $\mu_{Q_d}(k) = 1$ hay $\Pi_{Q_d\ r\ are\ P} = 1$. Do $\forall i \geq a$ thì $\mu_{Q_a}(i) = 1$ nên $\Pi_{Q_a\ r\ are\ P} = \max\limits_{k \leq i \leq n}\Big\{\min\Big(\mu_{Q_a}(i), \pi_{|r_P|}(i)\Big)\Big\} = \max\limits_{k \leq i \leq n}\Big\{\pi_{|r_P|}(i)\Big\} = \pi_{|r_P|}(k) = 1$.

Do đó theo Định lý 3.1, ta có $\Pi_{Q\ r\ are\ P} = 1$.

Nếu $k > b$ thì tương tự trên ta có $\Pi_{Q_a\ r\ are\ P} = 1$, và $\Pi_{Q_d\ r\ are\ P} = \mu_{Q_d}(k) = \mu_Q(k)$.

Do đó theo Định lý 3.1 ta có $\Pi_{Q\ r\ are\ P} = \mu_Q(k)$.

Tiếp theo ta chứng minh công thức (2).

Nếu $k \leq a$ ta có $\mu_{Q_a}(k) = \mu_Q(k)$ nên $N_{Q\ r\ areP} = \min\Big(\mu_Q(k), N_{Q_d\ r\ are\ P}\Big)$.

Nếu $k > a$ thì $\mu_{Q_a}(k) = 1$ nên $N_{Q\ r\ are\ P} = N_{Q_d\ r\ are\ P}$. ∎

Dựa vào Hệ quả 3.1 ta xây dựng thuật toán kiểm tra độ tin cậy của luật tổng kết dữ liệu dạng "Q r are P", với Q là lượng từ unimodal. Tương tự với thuật toán đã trình bày trong [1] cho lượng từ đơn điệu tăng, trong thuật toán này chúng tôi sử dụng 4 mảng PL, PU, NL, NU để lưu các cận dưới và cận trên của độ đo khả năng và độ đo cần thiết tại từng bước tính toán với mục đích dừng ngay thuật toán khi có thể kết luận được mà không cần phải duyệt hết tất cả các bộ của quan hệ r. Các mảng trên được cập nhật lại từng bước và thứ tự các phần tử luôn theo thứ tự tăng.

Thuật toán 3.1. Kiểm tra độ tin cậy của luật "Q r are P", với Q là lượng từ unimodal.

Input: r là một quan hệ với dữ liệu mờ

Lượng từ unimodal Q được phân tích thành $Q = Q_a \cap Q_d$

Tân từ P trên các bộ của quan hệ r

Ngưỡng xác định độ tin cậy cho độ đo khả năng và cần thiết, tương ứng α, β.

Output: *True* nếu $\Pi_{Q\ r\ are\ P} \geq \alpha$ và $N_{Q\ r\ are\ P} \geq \beta$

False trong trường hợp ngược lại.

Format: SatRule1(Q, r, P, α, β **)**

Method:

Var PL, PU, NL, NU : array[1..m] of Real;

Begin

/* Khởi tạo giá trị ban đầu cho các mảng */

```
        For i := 1 to m do
            PL[i]:=0; NL[i]:=0; PU[i]:=0; NU[i]:=0;
        EndFor
        For i := m + 1 to 2m − 1 do
            PL[i]:=μ_{Q_a}(i − m);
            NL[i]:=μ_{Q_d}(i − m);
        EndFor
        For i := 2m − n to 2m − 1 do
            PU[i]:=1; NU[i]:=1;
        EndFor
        For i := m − n + 1 to 2m − n − 1 do
            PU[i]:=μ_{Q_a}(i − m + n);
            NU[i]:=μ_{Q_d}(i − m + n);
        EndFor
        k := 0;
    /* Duyệt quan hệ r và tính toán */
        For i := 1 to n do
            If k + n − i < a then
                If (PL[m] ≥ α) and (μ_Q(k) ≥ β) and (NL[m]≥ β) then Return True;
                If (PU[m] < α) or (μ_Q(k + n − i) < β) or (NL[m]< β) then Return False;
            Else
                If (k ≥ a) then
                    If (k + n − i ≤ b) and (NL[m] ≥ β) then Return True;
                    If (NU[m] < β) then Return False;
                    If (k > b) then
                        If (μ_Q(k + n − i) ≥ α) and (NL[m] ≥ β) then Return True;
                        If (μ_Q(k) < α) or (NU[m] < β) then Return False;
                    EndIf
                EndIf
            EndIf
            x := μ_P(t_i);
            If (x = 1) then k := k + 1;
            Insert0(x,PL);
            Insert0(1 − x,NL);
            Insert1(x,PU);
            Insert1(1 − x,NU);
        EndFor
        Return (min(PL[m], μ_{Q_d}(k)) ≥ α) and (min(μ_{Q_a}(k), NL[m]) ≥ β);
    End
```

Chú ý:

Thủ tục Insert0(x,M) chèn x vào mảng M đã được sắp tăng, mỗi khi thêm thì bỏ đi một số 0 trong mảng. Thủ tục Insert1(x,M) thực hiện tương tự nhưng khi thêm thì bỏ đi một số 1 trong mảng.

Định lý 3.2. *Thuật toán 3.1 là đúng đắn và có độ phức tạp là $O(n^2)$, với $n = |r|$.*

Chứng minh Ký hiệu $PL[m]_i, PU[m]_i, NL[m]_i, NU[m]_i, k_i$ là giá trị tương ứng của các biến $PL[m]$, $PU[m]$, $NL[m]$, $NU[m]$, k tại bước thứ i của quá trình tính toán.

Trong [1] đã chứng minh được rằng $\forall i, 1 \leq i \leq n$ ta có

$$PL[m]_i \leq \Pi_{Q_a\ r\ are\ P} \leq PU[m]_i,\ NL[m]_i \leq N_{Q_d\ r\ are\ P} \leq NU[m]_i,$$

$$PL[m]_n = PU[m]_n = \Pi_{Q_a\ r\ are\ P},\ NL[m]_n = NU[m]_n = N_{Q_d\ r\ are\ P}.$$

Ta có $\forall\ 1 \leq i \leq n$ thì $k_i \leq k \leq k_i + n - i$. Tại bước thứ i, ta xét các trường hợp sau:

Trường hợp 1: $k_i + n - i < a$

Khi đó $k < a$, theo hệ quả 3.1 thì

$$\Pi_{Q\ r\ are\ P} = \Pi_{Q_a\ r\ are\ P} \text{ và } N_{Q\ r\ are\ P} = \min(\mu_Q(k), N_{Q_d\ r\ are\ P}).$$

Trong trường hợp này, nếu $PL[m]_i \geq \alpha$ thì $\Pi_{Q_a\ r\ are\ P} \geq \alpha$, hay $\Pi_{Q\ r\ are\ P} \geq \alpha$.

Nếu $PU[m]_i < \alpha$ thì $\Pi_{Q_a\ r\ are\ P} < \alpha$, hay $\Pi_{Q\ r\ are\ P} < \alpha$.

Tương tự, nếu $\mu_Q(k_i) \geq \beta$ thì $\mu_Q(k) \geq \beta$ (do $a > k \geq k_i$ và Q tăng trong đoạn $[0, a]$).

Nếu $NL[m]_i \geq \beta$ thì $N_{Q_d\ r\ are\ P} \geq \beta$. Từ đó $N_{Q\ r\ are\ P} = \min(\mu_Q(k), N_{Q_d\ r\ are\ P}) \geq \beta$.

Nếu $\mu_Q(k_i + n - i) < \beta$ thì $\mu_Q(k) < \beta$ (vì $k_i + n - i > k$, và Q tăng trong $[0, a]$).

Nếu $NU[m]_i < \beta$ thì $N_{Q_d\ r\ are\ P} < \beta$. Do đó $\Pi_{Q\ r\ areP} = \min(\mu_Q(k), N_{Q_d\ r\ are\ P}) < \beta$.

Trường hợp 2: $k_i + n - i \geq a$

Xét trường hợp con $k_i \geq a$. Khi đó $k \geq k_i \geq a$.

Trong trường hợp này, nếu $k_i + n - i \leq b$ thì $k \leq b$ nên $\Pi_{Q\ r\ are\ P} \geq \alpha$.

Do $k \geq a$ nên $N_{Q_d\ r\ are\ P} = N_{Q\ r\ are\ P}$.

Khi đó nếu $NL[m]_i \geq \beta$ thì $N_{Q_d\ r\ are\ P} \geq \beta$, hay $N_{Q\ r\ are\ P} \geq \beta$.

Ngược lại, nếu $NU[m]_i < \beta$ thì $N_{Q_d\ r\ are\ P} < \beta$, hay $N_{Q\ r\ are\ P} < \beta$.

Trường hợp 3: $k_i > b$

Trong trường hợp này $k > b$ nên $\Pi_{Q\ r\ are\ P} = \mu_Q(k)$, và $N_{Q\ r\ are\ P} = N_{Q_d\ r\ are\ P}$.

Khi đó, nếu $\mu_Q(k_i + n - i) \geq \alpha$ thì $\mu_Q(k) \geq \alpha$. Suy ra $\Pi_{Q\ r\ are\ P} \geq \alpha$.

Ngược lại, nếu $\mu_Q(k_i) < \alpha$ thì $\mu_Q(k) < \mu_Q(k_i) < \alpha$. Từ đó suy ra $\Pi_{Q\ r\ are\ P} < \alpha$.

Trong trường hợp này vì $N_{Q\ r\ are\ P} = N_{Q_d\ r\ are\ P}$ nên nếu $NL[m]_i \geq \beta$ thì $N_{Q\ r\ are\ P} \geq \beta$, và nếu $NU[m]_i < \beta$ thì $N_{Q\ r\ are\ P} < \beta$.

Đánh giá độ phức tạp:

Ta xem việc tính độ thoả tân từ P của bộ $t_i \in r$ ($\mu_P(t_i)$) có thời gian tính toán là hằng. Ta đánh giá độ phức tạp thuật toán trong trường hợp xấu nhất, tức là vòng lặp **For** duyệt hết tất cả các bộ của quan hệ r. Mỗi lần lặp thực hiện 4 thủ tục Insert, mỗi thủ tục thực hiện không quá $2n$ thao tác so sánh và gán. Số thao tác nhiều nhất của cả vòng lặp là $8n^2$. Do đó độ phức tạp của thuật toán là $O(n^2)$. ■

Ưu điểm của thuật toán là chỉ duyệt qua các bộ của quan hệ r không quá 1 lần, đồng thời những điều kiện dừng giúp cho thuật toán kết thúc sớm hơn trong các trường hợp cụ thể. Điều này đặc biệt quan trọng đối với những cơ sở dữ liệu lớn.

3.2. Dạng Q P_1 r *are* P_2

Trong [11] đã xây dựng công thức tính độ tin cậy cho các luật dạng này. Trong trường hợp Q là lượng từ tuyệt đối thì mệnh đề "$Q\ P_1\ r\ are\ P_2$" tương đương với "$Q\ r\ are\ P_1 \wedge P_2$", dạng luật này đã được giải quyết khá hoàn chỉnh trong phần trước. Trong phần này ta chỉ xét các luật dạng trên với Q là lượng từ tỷ lệ. Độ tin cậy của mệnh đề "$Q\ P_1\ r\ are\ P_2$" được đánh giá như sau:

$$\Pi_{Q\ P_1\ r\ are\ P_2} = \max_{k_1 \le i \le n} \Big\{ \min\Big(\Pi_{Q\ r^i\ are\ P_1 \wedge P_2}, \pi_{|r_{P_1}|}(i) \Big) \Big\} \tag{3.7}$$

$$N_{Q\ P_1\ r\ are\ P_2} = \min_{k_1 \le i \le n} \Big\{ \max\Big(N_{Q\ r^i\ are\ P_1 \wedge P_2}, 1 - \pi_{|r_{P_1}|}(i) \Big) \Big\} \tag{3.8}$$

với $n = |r|$, $k_1 = |\ker(r_{P_1})|$, $r^i \subseteq r$ là i bộ của quan hệ r có độ thoả tân từ P_1 cao nhất.

Kết quả sau thể hiện mối liên hệ giữa độ tin cậy của các luật dạng này trong trường hợp Q là lượng từ unimodal.

Định lý 3.3. *Cho Q là lượng từ unimodal được phân tích thành giao của hai lượng từ đơn điệu $Q = Q_a \cap Q_d$. Khi đó, với mọi quan hệ mờ r, với mọi tân từ P_1, P_2 trên r, ta có:*

$$\Pi_{Q\ P_1\ r\ are\ P_2} = \min\Big(\Pi_{Q_a\ P_1\ r\ are\ P_2}, \Pi_{Q_d\ P_1\ r\ are\ P_2} \Big),$$

$$N_{Q\ P_1\ r\ are\ P_2} = \min\Big(N_{Q_a\ P_1\ r\ are\ P_2}, N_{Q_d\ P_1\ r\ are\ P_2} \Big).$$

Chứng minh. Ký hiệu $k_1 = |\ker(r_{P_1})|$, $k_2 = |\ker(r_{P_1 \wedge P_2})|$. Trước hết ta chứng minh

$$\Pi_{Q\ P_1\ r\ are\ P_2} = \min\Big(\Pi_{Q_a\ P_1\ r\ are\ P_2}, \Pi_{Q_d\ P_1\ r\ are\ P_2} \Big) \tag{*}$$

Dựa vào Hệ quả 3.1, ta xét các trường hợp sau:

Trường hợp 1: $k_2 < a$

Khi đó $\forall\ k_1 \le i \le n$ $\Pi_{Q\ r^i\ are\ P_1 \wedge P_2} = \Pi_{Q_a\ r^i\ are\ P_1 \wedge P_2}$ nên $\Pi_{Q\ P_1\ r\ are\ P_2} = \Pi_{Q_a\ P_1\ r\ are\ P_2}$.

Trong trường hợp này $\Pi_{Q_a\ r^i\ are\ P_1 \wedge P_2} = \mu_{Q_d}(k_2) = 1$. Do đó công thức (*) đúng.

Trường hợp 2: $a \le k_2 \le b$

Khi đó theo Hệ quả 3.1, ta có $\Pi_{Q_a\ r^i\ are\ P_1 \wedge P_2} = 1$. Suy ra

$$\Pi_{Q\ P_1\ r\ are\ P_2} = \max_{k_1 \le i \le n} \Big\{ \pi_{|r_{P_1}|}(i) \Big\} = 1.$$

Dễ thấy $\Pi_{Q_a\ r^i\ are\ P_1 \wedge P_2} = 1$ nên $\Pi_{Q_a\ P_1\ r\ are\ P_2} = 1$, và $\Pi_{Q_d\ r^i\ are\ P_1 \wedge P_2} = 1$.

Suy ra $\Pi_{Q_d\ P_1\ r\ are\ P_2} = 1$. Do đó công thức (*) đúng.

Trường hợp 3: $k_2 > b$

Theo Hệ quả 3.1 ta có $\Pi_{Q\ r^i\ are\ P_1 \wedge P_2} = \mu_Q(k_2) = \Pi_{Q_d\ r^i\ are\ P_1 \wedge P_2}$.

Từ đó suy ra $\Pi_{Q\ P_1\ r\ are\ P_2} = \Pi_{Q_d\ P_1\ r\ are\ P_2}$.

Trong trường hợp này ta cũng có $\Pi_{Q_a\ r^i\ are\ P_1 \wedge P_2} = 1$. Do đó công thức (*) đúng.

Tiếp theo ta chứng minh

$$N_{Q\ P_1\ r\ are\ P_2} = \min\Big(N_{Q_a\ P_1\ r\ are\ P_2}, N_{Q_d\ P_1\ r\ are\ P_2} \Big) \tag{**}$$

Với mọi i, $k_1 \le i \le n$, theo Định lý 3.1, ta có

$$N_{Q\ r^i\ are\ P_1 \wedge P_2} = \min\Big(N_{Q_a\ r^i\ are\ P_1 \wedge P_2}, N_{Q_d\ r^i\ are\ P_1 \wedge P_2} \Big).$$

Do đó

$$\begin{aligned} N_{Q\ P_1\ r\ are\ P_2} &= \min_{k_1 \le i \le n} \Big\{ \max\Big(N_{Q\ r^i\ are\ P_1 \wedge P_2}, 1 - \pi_{|r_{P_1}|}(i) \Big) \Big\} \\ &= \min_{k_1 \le i \le n} \Big\{ \max\Big(\min(N_{Q_a\ r^i\ are\ P_1 \wedge P_2}, N_{Q_d\ r^i\ are\ P_1 \wedge P_2}), 1 - \pi_{|r_{P_1}|}(i) \Big) \Big\} \\ &= \min_{k_1 \le i \le n} \Big\{ \min\Big(\max\Big(N_{Q_a\ r^i\ are\ P_1 \wedge P_2}, 1 - \pi_{|r_{P_1}|}(i) \Big), \max\Big(N_{Q_d\ r^i\ are\ P_1 \wedge P_2}, 1 - \pi_{|r_{P_1}|}(i) \Big) \Big) \Big\} \\ &= \min\Big(N_{Q_a\ P_1\ r\ are\ P_2}, N_{Q_d\ P_1\ r\ are\ P_2} \Big). \end{aligned}$$

Công thức (**) được chứng minh. ∎

Hệ quả 3.2. *Với giả thiết như Định lý 3.3, ta có*

$\Pi_{Q\ P_1\ r\ are\ P_2} \ge \alpha$ *khi và chỉ khi* $\Pi_{Q_a\ P_1\ r\ are\ P_2} \ge \alpha$ *và* $\Pi_{Q_d\ P_1\ r\ are\ P_2} \ge \alpha$,

$N_{Q\ P_1\ r\ are\ P_2} \ge \beta$ *khi và chỉ khi* $N_{Q_a\ P_1\ r\ are\ P_2} \ge \beta$ *và* $N_{Q_d\ P_1\ r\ are\ P_2} \ge \beta$.

Chứng minh. Dễ dàng. ∎

Từ hệ quả trên, để kiểm tra độ tin cậy các luật dạng "$Q\ P_1\ r\ are\ P_2$", với Q là lượng từ unimodal ta có thể kiểm tra độ tin cậy luật tương ứng qua các lượng từ đơn điệu. Trong phần sau trình bày thuật toán kiểm tra luật "$Q\ P_1\ r\ are\ P_2$" với Q là lượng từ đơn điệu. Trước hết ta đưa ra các công thức tính độ tin cậy dựa vào đặc trưng của lượng từ đơn điệu.

Từ công thức (3.7), (3.8) ta xét các trường hợp cụ thể sau:

Trường hợp Q là lượng từ đơn điệu tăng:

Theo công thức (3.7), để tính $\Pi_{Q\ P_1\ r\ are\ P_2}$ ta phải tính $\Pi_{Q\ r^i\ are\ P_1 \wedge P_2}$, với $k_1 \le i \le n$.

Theo công thức (3.3), ta có

$\Pi_{Q\ r^i\ are\ P_1 \wedge P_2} = \text{med}\ \Big(\mu_Q(1/i), \cdots, \mu_Q((i-1)/i), \mu_{P_1 \wedge P_2}(t_1), \cdots, \mu_{P_1 \wedge P_2}(t_i) \Big)$, với $r^i = \{t_1, t_2, \cdots, t_i\}$ là i bộ của r có độ thoả tân từ P_1 cao nhất.

Do Q là lượng từ đơn điệu tăng nên theo (3.4), ta có $N_{Q\ r^i\ are\ P_1 \wedge P_2} = \mu_Q(k_2/i)$, với $k_2 = |\ker(r_{P_1 \wedge P_2})|$.

Vì $\mu_Q(k_2/i)$ là hàm giảm theo i và $1 - \pi_{|r_{P_1}|}(i)$ là hàm tăng theo i nên tương tự công thức (3.6), ta có:

$$N_{Q\ P_1\ r\ are\ P_2} = \text{med}\Big(\mu_Q(k_2/k_1), \cdots, \mu_Q(k_2/n), 1 - \mu_{P_1}(t_{k+1}), \cdots, 1 - \mu_{P_1}(t_n) \Big),$$

với giả thiết $r = \{t_1, t_2, \cdots, t_n\}$ thoả $\mu_{P_1}(t_1) \ge \mu_{P_1}(t_2) \ge \cdots \ge \mu_{P_1}(t_n)$.

Trường hợp Q là lượng từ đơn điệu giảm:

Tương tự trên, ta có kết quả trong trường hợp Q là lượng từ đơn điệu giảm:

$$\Pi_{Q\ P_1\ r\ are\ P_2} = \text{med}\Big(\mu_Q(k_2/k_1), \cdots, \mu_Q(k_2/n), \mu_{P_1}(t_{k+1}), \cdots, \mu_{P_1}(t_n) \Big),$$

và $N_{Q\ P_1\ r\ are\ P_2}$ được tính theo công thức (3.8), với $N_{Q\ r^i\ are\ P_1 \wedge P_2}$ được tính theo công thức:

$$N_{Q\ r^i\ are\ P_1 \wedge P_2} = \text{med}\Big(\mu_Q(1/i), \cdots, \mu_Q((i-1)/i), 1 - \mu_{P_1 \wedge P_2}(t_1), \cdots, 1 - \mu_{P_1 \wedge P_2}(t_i) \Big).$$

Từ những kết quả tính toán trên, thuật toán dưới đây cho phép kiểm tra độ tin cậy của luật tổng kết dữ liệu có dạng "$Q\ P_1\ r\ are\ P_2$", với lượng từ Q là đơn điệu tăng. Trong thuật toán này dùng mảng M các bản ghi với hai thành phần μ_1, μ_2, tương ứng chứa độ tin cậy của các tân từ P_1, $P_1 \wedge P_2$ cho mỗi bộ của quan hệ r. Ngoài ra, thuật toán dùng 2 mảng NL, NU với ý nghĩa như trong Thuật toán 3.1.

Thuật toán 3.2. Kiểm tra độ tin cậy của luật "$Q\ P_1\ r\ are\ P_2$", với Q là lượng từ đơn điệu tăng.

Input: r là một quan hệ với dữ liệu mờ
Lượng từ đơn điệu tăng Q
Tân từ P_1, P_2 trên các bộ của quan hệ r
Ngưỡng xác định độ tin cậy cho độ đo khả năng và cần thiết, tương ứng α, β.

Output: $True$ nếu $\Pi_{Q\ P_1\ r\ are\ P_2} \geq \alpha$ và $N_{Q\ P_1\ r\ are\ P_2} \geq \beta$
$False$ trong trường hợp ngược lại.

Format: SatRule2(Q**,** r**,** P_1**,** P_2**,** α, β **)**

Method:

Var

M : Array[1..n] of Record μ_1, μ_2: Real; End;
NL, NU : array[1..n] of Real;

Begin

/* Tính độ thoả tân từ P_1, P_2, $P_1 \wedge P_2$ của các bộ trong r */
$k_1 := 0$; $k_2 := 0$;
For $i := 1$ **to** n **do**
$x := \mu_{P_1}(t_i)$;
$y := \mu_{P_2}(t_i)$;
$z := \min(x, y)$;
If $x = 1$ **then** $k_1 := k_1 + 1$;
If $z = 1$ **then** $k_2 := k_2 + 1$;
Insert(x, z, M);
EndFor

/* Khởi tạo giá trị ban đầu cho các mảng NL, NU */
$h := n - k_1$;
For $i := 1$ **to** h **do** NL[i]:=0;
For $i := h + 1$ **to** $2h - 1$ **do** NL[i]:=$\mu_Q(k_2/(2n - i - 1))$;
For $i := 1$ **to** $h - 1$ **do** NU[i]:=$\mu_Q(k_2/(n - 1 - i))$;
For $i := h$ **to** $2h - 1$ **do** NU[i]:=1;
P:=0; N:=0;

/* Tính toán và kiểm tra độ tin cậy */
For $i := 1$ **to** n **do**
If $(M[i].\mu_1 < \alpha)$ or $(MU[h] < \beta)$ **then Return** $False$;
$P_i :=$ med $\Big(\mu_Q(1/i), \cdots, \mu_Q((i-1)/i), M[1].\mu_2, \cdots, M[i].\mu_2\Big)$;
Insert0$(1 - M[i].\mu_1, NL)$;
Insert1$(1 - M[i].\mu_1, NU)$;
$P := \max\Big(P, \min(P_i, M[i].\mu_1)\Big)$;
If $(P \geq \alpha)$ and $(NL[h] \geq \beta)$ **then Return** $True$;

EndFor
Return $(P \geq \alpha)$ and $(NL[h]) \geq \beta)$;
End

Chú ý:

Thủ tục Insert0(x,M) và Insert1(x,M) như trong Thuật toán 3.1. Thủ tục Insert(x, z, M) chèn x, z vào mảng M đã được sắp giảm theo thành phần μ_1.

Định lý 3.4. *Thuật toán 3.2 là đúng đắn và có độ phức tạp là $O(n^3)$, với $n = |r|$.*

Chứng minh. Với những công thức tính $\Pi_{Q\ P_1\ r\ are\ P_2}$ và $N_{Q\ P_1\ r\ are\ P_2}$, trong trường hợp Q là lượng từ đơn điệu tăng đã xây dựng, dễ thấy kết quả tính toán của thuật toán cho ta $P = \Pi_{Q\ P_1\ r\ are\ P_2}$, và $N = N_{Q\ P_1\ r\ are\ P_2}$.

Thuật toán dừng khi xảy ra điều kiện $NU[h] < \beta$ hoặc $NL[h] \geq \beta$, tương tự như trong Thuật toán 3.1.

Tại mỗi bước lặp tính giá trị P ta luôn có $P \leq \Pi_{Q\ P_1\ r\ are\ P_2}$, nên tại bước nào đó mà ta có $P \geq \alpha$ thì $\Pi_{Q\ P_1\ r\ are\ P_2} \geq \alpha$. Trong trường hợp này kết hợp với điều kiện $NL[h] \geq \beta$ thuật toán dừng và trả về kết quả *True* là đúng.

Vì giá trị μ_{P_1} của các bộ trong r được sắp theo thứ tự giảm trong mảng M, mà tại bước thứ $i-1$ nếu $P < \alpha$ và $M[i].\mu_1 < \alpha$ thì từ phép gán $P := \max(P, \min(P_i, M[i].\mu_1))$ ta thấy rằng bắt đầu từ bước thứ i trở đi ta luôn có $P < \alpha$. Do đó $\Pi_{Q\ P_1\ r\ are\ P_2} < \alpha$. Thuật toán dừng tại bước thứ i và kết luận *False* là đúng.

Đánh giá độ phức tạp:

Đoạn chương trình tính μ_{P_1}, μ_{P_2} lặp n lần thủ tục Insert. Thực chất đây là đoạn chương trình sắp xếp bằng chèn trực tiếp trên mảng M nên độ phức tạp là $O(n^2)$.

Đoạn chương trình tính $\Pi_{Q\ P_1\ r\ are\ P_2}$ và $N_{Q\ P_1\ r\ are\ P_2}$ trong trường hợp xấu nhất lặp $n+1$ lần (trường hợp $k_1 = 0$). Trong vòng lặp này, công việc lặp có độ phức tạp lớn nhất là tính $P_i := \text{med}\left(\mu_Q(1/i), \cdots, \mu_Q((i-1)/i), M[1].\mu_2, \cdots, M[i].\mu_2\right)$, có độ phức tạp là $O(n^2)$. Do đó đoạn chương trình này có độ phức tạp là $O(n^3)$. Đây cũng là độ phức tạp của thuật toán.

■

3.3. Dạng $Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r$

Trong [11], độ tin cậy của luật dạng này được đánh giá qua dạng tương đương "$Q_1\ P_1\ r\ are\ P$", với P là tân từ áp dụng cho mỗi bộ t_i của r và được đánh giá bởi công thức: $\mu_P(t_i) = \Pi_{Q_2\ P_2\ r\ are\ t_i \circ \theta}$, với $t_i \circ \theta$ là phép hợp thành một giá trị mờ của bộ t_i với phép so sánh mờ θ (xem 2.1).

Định lý 3.5. *Cho Q_1 là lượng từ unimodal được phân tích thành giao của hai lượng từ đơn điệu $Q_1 = Q_{1a} \cap Q_{1d}$. Khi đó với mọi quan hệ r, với mọi lượng từ Q_2, với mọi tân từ P_1, P_2 trên quan hệ r, với mọi toán tử so sánh θ, ta có*

$$\Pi_{Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r} = \min\Big(\Pi_{Q_{1a}\ P_1\ r\ \theta\ Q_2\ P_2\ r}, \Pi_{Q_{1d}\ P_1\ r\ \theta\ Q_2\ P_2\ r}\Big),$$
$$N_{Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r} = \min\Big(N_{Q_{1a}\ P_1\ r\ \theta\ Q_2\ P_2\ r}, N_{Q_{1d}\ P_1\ r\ \theta\ Q_2\ P_2\ r}\Big).$$

Chứng minh. Dễ dàng chứng minh nhờ kết quả của Định lý 3.3. ■

Hệ quả 3.3. *Với giả thiết như Định lý 3.5, ta có*

$\Pi_{Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r} \geq \alpha$ *khi và chỉ khi* $\Pi_{Q_{1a}\ P_1\ r\ \theta\ Q_2\ P_2\ r} \geq \alpha$ *và* $\Pi_{Q_{1d}\ P_1\ r\ \theta\ Q_2\ P_2\ r} \geq \alpha$,
$N_{Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r} \geq \beta$ *khi và chỉ khi* $N_{Q_{1a}\ P_1\ r\ \theta\ Q_2\ P_2\ r} \geq \beta$ *và* $N_{Q_{1d}\ P_1\ r\ \theta\ Q_2\ P_2\ r} \geq \beta$.

Chứng minh. Dễ dàng. ∎

Thuật toán dưới đây cho phép kiểm tra luật tổng kết dữ liệu có dạng "$Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r$" cho lượng từ Q_1 là đơn điệu tăng. Thuật toán này tương tự Thuật toán 3.2.

Thuật toán 3.3. Kiểm tra độ tin cậy của luật "$Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r$", với Q_1 là lượng từ đơn điệu tăng.

Input: r là một quan hệ với dữ liệu mờ
Q_1, Q_2 là các lượng từ, trong đó Q_1 là lượng từ đơn điệu tăng
Tân từ P_1, P_2 trên các bộ của quan hệ r
Ngưỡng xác định độ tin cậy cho độ đo khả năng và cần thiết, tương ứng α, β.

Output: *True* nếu $\Pi_{Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r} \geq \alpha$ và $N_{Q_1\ P_1\ r\ \theta\ Q_2\ P_2\ r} \geq \beta$
False trong trường hợp ngược lại.

Format: SatRule3(Q_1**,** Q_2**,** r**,** P_1**,** P_2**,** α, β **)**

Method:

Var
M : Array[1..n] of Record μ_1, μ_2: Real; End;
NL, NU : array[1..n] of Real;

Begin
/* Tính độ thoả tân từ P_1, P của các bộ trong r */
$k_1 := 0$; $k_2 := 0$;
For $i := 1$ **to** n **do**
$x := \mu_{P_1}(t_i)$;
$y := \Pi_{Q_2\ P_2\ r\ are\ t_i \circ \theta}$; /* Thay đổi so với thuật toán 3.2 */
$z := \min(x, y)$;
If $x = 1$ **then** $k_1 := k_1 + 1$;
If $z = 1$ **then** $k_2 := k_2 + 1$;
Insert(x, z, M);
EndFor
$\cdots$ /* Như thuật toán 3.2 */
End

Định lý 3.6. *Thuật toán 3.3 là đúng đắn và có độ phức tạp là O(*n^4*) với* $n = |r|$.

Chứng minh. Tính đúng đắn của thuật toán đã được kiểm chứng qua tính đúng đắn của thuật toán 3.2. Theo định lý 3.4, độ phức tạp của thuật toán tính $\Pi_{Q_2\ P_2\ r\ are\ t_i \circ \theta}$ là O(n^3). Do đó dễ thấy độ phức tạp của thuật toán là O(n^4). ∎

4. KẾT LUẬN

Những kết quả tính toán độ tin cậy của một số dạng luật tổng kết dữ liệu được trình bày trong bài báo là cơ sở tốt cho việc xây dựng các thuật toán tính toán và kiểm tra độ tin cậy của các luật với ngưỡng cho trước. Những thuật toán sử dụng kỹ thuật heuristics tạo những điểm dừng hợp lý, giúp cho việc kết thúc thuật toán được sớm hơn. Điều này có ý nghĩa trong việc khai thác các luật từ những cơ sở dữ liệu lớn. Mặc dù đã có những cải tiến nhưng một số thuật toán vẫn có độ phức tạp khá lớn và là một trong những trở ngại khi khai thác các luật từ dữ liệu. Các thuật toán này chúng tôi tiếp tục nghiên cứu và cải tiến để tốt hơn.

Lời cảm ơn

Tác giả xin chân thành cảm ơn PGS.TS. Hồ Thuần đã đóng góp những ý kiến quý báu trong quá trình hoàn thành bài báo này.

TÀI LIỆU THAM KHẢO

[1] P. Bosc, L. Lietard, and O. Pivert, Quantified statements and Database Fuzzy querying, *Fuzziness in Database Management Systems*, Bosc P., Kacprzyk J. eds., Physica Verlag, (1995) 275–308.

[2] J. C. Cubero, J. M. Medina, O. Pons, M. A. Vila, Data summarization in relational databases through fuzzy dependencies, *Information Sciences* **121** (1999) 233–270.

[3] D. Dubois, H. Prade, *Possibility Theory: An Approach to Computerized Processing of Uncertainty*, Plenum Press, New York, 1988.

[4] D. Dubois, H. Prade, Fuzzy cardinality and the modeling of imprecise quantification, *Fuzzy Sets and Systems* **16** (1985) 199 – 230.

[5] J. Kacprzyk, A. Ziólkowski, Database Queries with Fuzzy Linguistic Quantifiers, *IEEE Transactions on Systems, Man and Cybernetics* **16** (3) (1986) 474 – 479.

[6] D. Rasmussen, R. R Yager, SummarySQL - A Fuzzy Tool for Data Mining, *Intelligent Data Analysis*, **1**(1) (1997).

[7] D. Rasmussen, R. R. Yager, Finding fuzzy and gradual functional dependencies with SummarySQL, *Fuzzy Sets and Systems* **106** (1999) 134–142.

[8] L. T. Vuong, H. Thuan, A relational databases extended by application of Fuzzy set theory and linguistic variables, *Computers and Artifical Intelligence* **8** (2) (1989) 153–168.

[9] H. Thuan, T. T. Thanh, On the Functional Dependencies and Multivalued Dependencies in Fuzzy relational databases, *Journal of Computer science and Cybernetics*, T17, **2**(2001), 13-19.

[10] H. Thuan, T. T. Thanh, Fuzzy Functional Dependencies With Linguistic Quantifiers, *Journal of Computer science and Cybernetics* **18** (2) (2002) 97–108.

[11] T. T. Thành, Một số mở rộng tổng kết dữ liệu trên cơ sở dữ liệu quan hệ mờ, *Tạp chí Tin học và Điều khiển học* **19** (1) (2003).

[12] R. R. Yager, Fuzzy Summaries in Database Mining, *Proceedings of the 11th Conference on Artifical Intelligence for Applications*, Los Angeles, 1995 (265–269).

[13] L. Zadeh, Fuzzy sets as a Basis for Theory of Possibility, *Fuzzy Sets and Systems* **13** (1978) 3–28.

[14] L. A. Zadeh, A Computational Approach to Fuzzy Quantifiers in Natural Languages, *Computers and Mathematics with Applications* **9** (1983) 149–184.

Nhận bài ngày 06 - 5 - 2003