Tạp chí Tin học và Điều khiển học, T. 16, S. 4 (2000), 79-84

THUẬT TOÁN TÌM BAO ĐÓNG CỦA TẬP SỰ KIỆN VÀ LOẠI BỎ LUẬT DƯ THỪA CỦA TẬP LUẬT TRONG HỆ LUẬT CỦA HỆ CHUYÊN GIA

LÊ HẢI KHÔI

Abstract. In this paper we give algorithms for finding the closure of the facts set and for removing redundant rules of the rules set in the rule-based system of the export system.

Tóm tắt. Mục đích của bài báo là cung cấp một số thuật toán liên quan đến việc tìm bao đóng của tập sự kiện và loại bỏ sự dư thừa của hệ luật trong hệ chuyên gia. Tính đúng đắn của thuật toán được chứng minh chặt chẽ dưới góc độ toán học.

1. MỞ ĐẦU

Hệ chuyên gia là một thành tựu của công nghệ tri thức. Mục tiêu chính của hệ chuyên gia là mô phỏng các hoạt động của người chuyên gia trên máy tính. Bản chất của hệ chuyên gia là một hệ phần mềm thông minh dạy cho máy các hoạt động của người chuyên gia.

Hệ chuyên gia thông thường gồm năm thành phần chính sau: cơ sở tri thức, mô tơ suy diễn, giao diện người dùng, bộ giải thích và bộ thu nạp tri thức. Trong các thành phần này quan trọng nhất là cơ sở tri thức và mô tơ suy diễn. Có thể nói rằng "Hệ chuyên gia = Cơ sở tri thức + Mô tơ suy diễn".

Cơ sở tri thức được biểu diễn bằng nhiều phương pháp: phương pháp logic, phương pháp mạng ngữ nghĩa, phương pháp mô hình, phương pháp hệ luật, phương pháp thông qua khung, phương pháp bộ ba OAV (đối tượng - thuộc tính - giá trị), v.v.. Các phương pháp biểu diễn tri thức tiến hành mô tả các đối tượng, các sự kiện, các quan niệm và một phần các mối quan hệ giữa chúng với nhau. Mỗi phương pháp biểu diễn tri thức đều có những ưu điểm cũng như nhược điểm nhất định. Thông thường, để tiến hành biểu diễn tri thức người ta phân tách toàn bộ biểu diễn tri thức thành các bài toán nhỏ hơn, đối với mỗi bài toán nhỏ lựa chọn phương pháp thích hợp để tận dụng ưu điểm, rồi liên kết chúng lại với nhau.

Tuy vậy, trong các phương pháp biểu diễn tri thức thì phương pháp biểu diễn bằng hệ luật là phương pháp tương đối phổ biến, nhờ các ưu điểm sau:

- Cách biểu diễn trực quan và đơn giản.
- Có thể kiểm tra tính mâu thuẫn trong hệ luật.
- Có tính mô đun cao (có thể thêm hoặc bớt các luật mà không phụ thuộc vào các luật khác).
- Có thể xử lý mâu thuẫn và dư thừa.

Những kiến thức cơ sở về hệ chuyên gia và các phương pháp biểu diễn tri thức có thể tìm trong [1, 2, 4, 5].

Cấu trúc của bài báo như sau. Mục 2 dành cho việc trình bày các khái niệm cơ bản liên quan đến các hệ luật mà cần thiết cho các mục tiếp theo. Mục 3 đưa ra một thuật toán tìm bao đóng của tập sự kiện và chứng minh tính đúng của thuật toán bằng phương pháp qui nạp toán học. Mục 4 đề cập thuật toán loại bỏ luật dư thừa của tập luật và sự dư thừa của hệ luật. Tính đúng của thuật toán được chứng minh bằng phương pháp phản chứng. Cuối cùng, mục 5 nêu lên một số vấn đề mở.

2. CÁC KHÁI NIỆM CƠ BẢN

Người ta dùng hệ luật bao gồm các câu "nếu... thì..." để biểu diễn tri thức theo cấu trúc sau:

nếu ⟨điều kiện 1⟩, ⟨điều kiện 2⟩,... , ⟨điều kiện m⟩
thì ⟨kết luận 1⟩, ⟨kết luận 2⟩,... , ⟨kết luận n⟩. (dạng 1)

Trong hệ luật trên các điều kiện và kết luận được thể hiện tương đối tự do.

Chúng ta có thể hình thức hóa cao hơn để thể hiện toàn bộ tri thức trong một hệ luật. Cụ thể như sau.

Định nghĩa 2.1. Hệ luật, kí hiệu là $L = \langle F, R\rangle$, gồm hai thành phần $F = \{f_1, ..., f_p\}$ là *tập các sự kiện*, $R = \{r_1, ..., r_q\}$ là *tập các luật*.

Thông thường F là tập hợp bao gồm tất cả các sự kiện xuất hiện trong luật, nằm ở vế phải và vế trái của các luật, mỗi luật đó thể hiện bằng cú pháp

$$A \to B,$$

trong đó A và B là những biểu thức bao gồm các sự kiện nối với nhau bằng các phép "và" ($\wedge$), "hoặc" ($\vee$), "phủ định" ($\neg$). Trong luật này ta hiểu là "nếu A đúng thì có B".

Dễ dàng thấy rằng một khi có luật "nếu $P_1 \vee P_2$ thì Q", chúng ta luôn có thể tách luật này thành hai luật "nếu P_1 thì Q" và "nếu P_2 thì Q". Hơn thế nữa, theo các qui tắc biến đổi của Vương Hạo, chúng ta luôn có thể chuyển đổi tương đương một hệ luật bất kỳ thành hệ luật chỉ bao gồm các luật dạng

$$P_1 \wedge P_2 \wedge \cdots \wedge P_n \to Q,$$

ở đây $P_1, P_2, \ldots, P_n$ và Q là các sự kiện. Điều này có nghĩa là "nếu tất cả các P_i ($i = 1, 2, \ldots, n$) là đúng thì ta có Q". Như vậy, chúng ta có một hệ luật gồm các luật với vế trái chỉ toàn là phép $\wedge$ và vế phải chỉ có một sự kiện.

Để đơn giản, chúng ta thay dấu $\wedge$ trong vế trái bằng dấu phảy (,) khi đó được luật dạng

$$P_1, P_2, \ldots, P_n \to Q.$$

Giả sử có hệ luật $L = \langle F, R\rangle$, trong đó $F = \{f_1, \ldots, f_p\}$ là tập các sự kiện, $R = \{r_1, \ldots, r_q\}$ là tập các luật. Kí hiệu F^* là tập các sự kiện $f \in F$ thỏa mãn đồng thời hai điều kiện:

(i) f có mặt ở vế trái,

(ii) f không có mặt ở vế phải,

trong tất cả các luật thuộc R. Tập F^* này được gọi là *tập các sự kiện gốc*.

Ví dụ 1. $L = \langle F, R\rangle$, với $F = \{a, b, c, h, k\}$ và $R = \{r_1, r_2\}$, trong đó "r_1 : nếu a, b thì h" và "r_2: nếu b, c thì k". Khi đó $F^* = \{a, b, c\}$.

Nếu kí hiệu F_0 là *tập các sự kiện ban đầu*, thì thông thường $F_0 \subseteq F^*$. Nói chung các điều kiện đối với F_0 tương đối tự do. Nếu từ F_0 suy diễn để tìm ra kết luận, thì suy diễn đó được gọi là *suy diễn tiến*. Còn nếu từ $F' \subseteq F$ ta suy về F'', mà tập F'' này là các điều kiện cho trước, thì suy diễn này được gọi là *suy diễn lùi*.

Trong việc biểu diễn tri thức bằng hệ luật còn có một loại hệ luật có cấu trúc như sau:

nếu ⟨điều kiện 1⟩, ⟨điều kiện 2⟩,... , ⟨điều kiện m⟩
thì ⟨thực hiện 1⟩, ⟨thực hiện 2⟩,... , ⟨thực hiện n⟩ (dạng 2)

trong đó các thực hiện có thể làm thay đổi các biến tham gia trong các điều kiện. Nói cách khác, các luật có tác động vào tập các sự kiện.

Ví dụ 2. Với hệ luật $L = \langle F, R\rangle$, trong đó $F = \{x = 5, y = 4\}$, $R = \{r\}$ và luật r được cho như sau: "r : nếu x lẻ, y chẵn, thì $x := x - 3$, $y := y + 2$". Khi đó $F_r = \{x = 2,\ y = 6\}$, ở đây $F_r := r(F)$ là kí hiệu của tập các sự kiện thu được từ F sau khi đã có tác động của luật r.

Chúng ta nói rằng luật r là *thích ứng* với tập sự kiện $F' \subseteq F$, nếu r thực hiện được với các sự kiện của F'. Trong trường hợp ngược lại, chúng ta nói rằng r *không thích ứng* với F'. Trong ví dụ 2 thì r thích ứng với F, nhưng không thích ứng với F_r.

Định nghĩa 2.2. Hệ luật $L = \langle F, R\rangle$ với $F = \{f_1, ..., f_p\}$ và $R = \{r_1, ..., r_q\}$, được gọi là *đơn điệu*, nếu với mọi cặp luật r_i và r_j ($i \neq j$), một khi chúng đã thích ứng với tập sự kiện $F' \subseteq F$ nào đó, thì sau khi áp dụng luật r_i cho F' để có F'_{r_i}, luật r_j cũng thích ứng với F'_{r_i} và đối với r_j cũng vậy.

Nếu không thỏa mãn điều kiện này thì L được gọi là hệ luật *không đơn điệu*.

Không sợ nhầm lẫn, chúng ta có thể kí hiệu $Left(r)$ là tập các sự kiện ở vế trái của luật r và $Right(r)$ là tập các sự kiện ở vế phải của luật r. Khi đó, với kí hiệu vừa nêu, có thể biểu diễn hệ luật đơn điệu như sau: cho "$r_1 : Left(r_1) \rightarrow Right(r_1)$" và "$r_2 : Left(r_2) \rightarrow Right(r_2)$", nếu $Left(r_1) \subseteq F'$, $Left(r_2) \subseteq F'$, thì ta có $Left(r_2) \subseteq F'_{r_1}$, $Left(r_1) \subseteq F'_{r_2}$. Thế thì tính không đơn điệu có thể hiểu là: tồn tại cặp (r_i, r_j) và $F' \subseteq F$ sao cho nếu r_i, r_j thích ứng với F', thì hoặc r_i không thích ứng với F'_{r_j} hoặc r_j không thích ứng với F'_{r_i}.

Định nghĩa 2.3. Hệ luật $L = \langle F, R \rangle$ với $F = \{f_1, ..., f_p\}$ và $R = \{r_1, ..., r_q\}$, được gọi là *giao hoán bộ phận*, nếu với mọi cặp luật r_i và r_j $(i \neq j)$, với tập sự kiện $F' \subseteq F$ bất kỳ, ta luôn có

$$r_i(r_j(F')) = r_j(r_i(F')).$$

Ở đây, $r(F')$ hiểu theo nghĩa: nếu r thích ứng với F' thì $r(F') = F'_r$, còn nếu r không thích ứng với F' thì $r(F') = F'$.

Ví dụ 3 (hệ luật đơn điệu, nhưng không giao hoán bộ phận). Xét hệ luật $L = \langle F, R \rangle$ với $F = \{x = 2,\ y = 8\}$ và $R = \{r_1, r_2\}$, trong đó:

"r_1: nếu $x =$ nguyên tố, $y =$ chẵn, thì $x := x + 2,\ y := y/2$",

"r_2: nếu $x =$ chẵn, $y =$ chẵn, thì $x := x + 3,\ y := y + 4$".

Khi đó, $r_1(F) = \{x = 4,\ y = 4\}$ và $r_2(r_1(F)) = \{x = 7,\ y = 8\}$, còn $r_2(F) = \{x = 5,\ y = 12\}$ và $r_1(r_2(F)) = \{x = 7,\ y = 6\}$. Như vậy, hệ luật là đơn điệu, nhưng không giao hoán bộ phận.

Ví dụ 4 (hệ luật giao hoán bộ phận, nhưng không đơn điệu). Xét hệ luật $L = \langle F, R \rangle$ với $F = \{x = 6,\ y = 3\}$ và $R = \{r_1, r_2\}$, trong đó:

"r_1: nếu $x =$ hợp số, $y =$ nguyên tố, thì $x := x + 3, y := y * 2$",

"r_2: nếu $x =$ chẵn, $y =$ lẻ, thì $x := x + x/2,\ y := y + 3$".

Khi đó, $r_1(F) = \{x = 9,\ y = 6\}$, hơn nữa do r_2 không thích ứng với F_{r_1}, nên $r_2(r_1(F)) = \{x = 9,\ y = 6\}$. Mặt khác $r_2(F) = \{x = 9,\ y = 6\}$ và do r_1 cũng không thích ứng với F_{r_2} nên $r_1(r_2(F)) = \{x = 9,\ y = 6\}$. Vậy là hệ luật này giao hoán bộ phận, nhưng không đơn điệu.

Định nghĩa 2.4. Hệ luật $L = \langle F, R \rangle$ được gọi là *giao hoán*, nếu hệ này đồng thời là đơn điệu và giao hoán bộ phận.

Dễ dàng nhận thấy hệ luật dạng 1 là một hệ luật giao hoán.

Từ đây trở đi, trong phạm vi bài báo này, chúng ta chỉ đề cập các hệ luật dạng 1. Ngoài ra, chúng ta sử dụng ký hiệu $\langle ... \rangle$ để chỉ dãy (tức là có thứ tự) các phần tử.

3. BAO ĐÓNG CỦA TẬP SỰ KIỆN VÀ CÁCH TÌM

Trong mục này, chúng ta đề cập việc tính bao đóng của một tập sự kiện. Giả sử có hệ luật $L = \langle F, R \rangle$ với $F = \{f_1, \dots, f_p\}$ và $R = \{r_1, \dots, r_q\}$, trong đó mọi luật $r \in R$ đều có dạng "r: nếu $P_1, P_2, \dots, P_l$ thì Q" và $F' \subseteq F$. *Bao đóng* của F' đối với R, kí hiệu là $(F'_R)^+$, hay đơn giản là ${F'_R}^+$, là tập thu được từ F' sau khi áp dụng tất cả các luật có thể có của R.

Dưới đây luôn giả thiết là các phép suy diễn không bị lặp (tức là không có chu trình).

Thuật toán 3.1. (tính ${F'_R}^+$)

Input: $L = \langle F, R \rangle$ với $F = \langle f_1, ..., f_p \rangle$, $R = \langle r_1, ..., r_q \rangle$ và $F' \subseteq F$.

Output: ${F'_R}^+$.

- Bước 0: đặt $K_0 = F'$;
- Bước i: nếu có luật $r \in R$ thỏa mãn điều kiện $Left(r) \subseteq K_{i-1}$ và $Right(r) \not\subseteq K_{i-1}$, thì đặt $K_i = K_{i-1} \cup Right(r)$.

- Quá trình được lặp lại cho đến khi $K_i = K_{i+1}$.
Lúc đó đặt ${F'_R}^+ = K_i$.

Định lý 3.2. *Thuật toán 3.1 là dừng và cho kết quả là bao đóng ${F'_R}^+$ của tập sự kiện $F' \subseteq F$.*

Chứng minh. Chúng ta sử dụng phương pháp qui nạp toán học.

Từ thuật toán suy ra rằng đến một chỉ số n nào đó, bắt đầu từ K_n, thì dừng: $K_0 \subset K_1 \subset \cdots \subset K_n = K_{n+1}$. Rõ ràng rằng n không thể lớn hơn q là số lượng các phần tử của tập R. Chúng ta chứng minh rằng ${F'_R}^+ = K_n$.

Trước hết nhận xét rằng bao hàm thức $K_n \subseteq {F'_R}^+$ là hiển nhiên, vì mọi K_i đều có được từ F' qua những tác động của các luật thuộc R.

Vấn đề còn lại là chứng minh ${F'_R}^+ \subseteq K_n$. Do $F' = K_0 \subset K_n$, nên chúng ta chỉ còn phải chứng minh rằng ${F'_R}^+ \setminus F' \subseteq K_n$ là xong.

Để ý rằng mỗi một sự kiện thuộc tập ${F'_R}^+ \setminus F'$ đều là kết quả của sự tác động vào F' của một dãy (hữu hạn) nào đó các luật (về nguyên tắc, có thể có nhiều dãy như thế), do đó chúng ta sẽ xem xét số các số hạng của dãy (hay còn gọi là độ dài của dãy). Chúng ta sẽ chứng minh bằng qui nạp theo $t \in N$ rằng bất kỳ sự kiện nào sinh ra bởi dãy có độ dài t đều thuộc K_n.

Không mất tính tổng quát của bài toán, chúng ta có thể giả thiết rằng sự tác động của các luật đều là thực sự, có nghĩa là mỗi một luật, sau khi tác động vào tập sự kiện nào đó đều sinh ra một sự kiện mới không thuộc tập sự kiện mà nó vừa tác động (nếu không như thế thì tác động của luật sẽ trở thành thừa). Trước khi bước vào chứng minh, chúng ta qui ước rằng luật r trong bước i của thuật toán sẽ được gọi là luật sinh ra tập K_i.

Với $t = 1$. Trong trường hợp này, tồn tại một luật nào đó $r \in R$, thích ứng với F' và $Right(r) = f \notin F'$. Có hai khả năng xảy ra:

- Luật r là một trong các luật sinh ra $K_1, \ldots, K_n$, chẳng hạn sinh ra K_m nào đó. Điều đó có nghĩa là $Letf(r) \subseteq K_{m-1}$ và $Right(r) = f \notin K_{m-1}$. Từ đó, theo thuật toán, chúng ta có $K_m = K_{m-1} \cup Right(r)$, suy ra $f \in K_m \subset K_n$.

- Luật r không thuộc tập các luật sinh ra $K_1, \ldots, K_n$. Thế thì, do bắt đầu từ K_n thì việc sinh thêm sự kiện mới dừng lại, nên $f = Right(r)$ phải thuộc K_n.

Vậy là với $t = 1$ bài toán đúng.

Bây giờ giả sử rằng mọi dãy luật có độ dài không vượt quá t, khi tác động vào F' đều cho kết quả là một sự kiện thuộc K_n. Chúng ta xét dãy có độ dài $t + 1$, chẳng hạn, $r_{\alpha_1}, \ldots, r_{\alpha_{t+1}}$. Ký hiệu $L_1, \ldots, L_{t+1}$ là tập các sự kiện do dãy này tác động vào F' sinh ra:

$$F' \subset L_1 \subset \cdots \subset L_t \subset L_{t+1}.$$

Xét tác động của t luật đầu $r_{\alpha_1}, \ldots, r_{\alpha_t}$, ta có: $Left(r_{\alpha_t}) \subseteq L_{t-1}$, $Right(r_{\alpha_t}) = g \notin L_{t-1}$ và $L_t = L_{t-1} \cup \{g\}$. Theo giả thiết qui nạp, $g \in K_n$ và dễ dàng thấy $L_t \subseteq K_n$.

Đối với $r_{\alpha_{t+1}}$ ta có: $Left(r_{\alpha_{t+1}}) \subseteq L_t$ và $Right(r_{\alpha_{t+1}}) = f \notin L_t$. Từ $L_t \subseteq K_n$ ta suy ra $Left(r_{\alpha_{t+1}}) \subseteq K_n$. Theo thuật toán, vì đến K_n là dừng, có nghĩa là $\forall r \in R$ mà $Left(r) \subseteq K_n$ thì $Right(r) \in K_n$, nên chúng ta có $Right(r_{\alpha_{t+1}}) = f \in K_n$.

Vậy, ${F'_R}^+ \setminus F' \subseteq K_n$, thuật toán được chứng minh.

Dễ dàng chứng minh kết quả sau.

Mệnh đề 3.3. *Thuật toán tính bao đóng nêu trên là thuật toán có độ phức tạp đa thức theo lực lượng của F và R.*

Nhận xét. 1) Việc tính ${F'_R}^+$ chỉ thực sự có ý nghĩa nếu như $F' \subseteq F^*$, ở đây F^* là tập các sự kiện chỉ có mặt ở vế trái của mọi luật thuộc R.

2) Trong lý thuyết cơ sở dữ liệu quan hệ cũng có thuật toán tìm bao đóng (của tập thuộc tính) [3], tuy nhiên suy diễn ở đó dựa trên hệ tiên đề Armstrong, hoàn toàn khác với suy diễn logic trong hệ luật của hệ chuyên gia.

4. LOẠI BỎ DƯ THỪA TRONG TẬP LUẬT VÀ HỆ LUẬT

Bây giờ chúng ta chuyển sang việc xử lý dư thừa trong hệ luật. Về mặt mô tả, sự dư thừa có thể hiểu là: một sự kiện hoặc một luật được gọi là dư thừa trong cơ sở tri thức hệ luật, nếu nó không ảnh hưởng đến toàn bộ quá trình suy diễn.

Về mặt toán học, sự dư thừa của tập luật có thể định nghĩa như sau. Xét hệ luật $L = \langle F, R\rangle$, F^* là tập các sự kiện chỉ tham gia trong vế trái mà không tham gia trong vế phải của các luật. Nếu có $r \in R$ sao cho ${F_R^*}^+ = (F^*_{R\setminus\{r\}})^+$, thì r được coi là *thừa* và chúng ta có thể loại bỏ luật này đi.

Trên cơ sở thuật toán tính bao đóng, chúng ta xây dựng thuật toán sau.

Thuật toán 4.1. (loại bỏ luật thừa)

Input: $L = \langle F, R\rangle$ với $F = \langle f_1, \ldots, f_p\rangle$ và $R = \langle r_1, \ldots, r_q\rangle$.

Output: R' thỏa mãn $R' \subseteq R$, $(F^*_{R'})^+ = {F_R^*}^+$ và $\forall r \in R' : R'' = R' \setminus \{r\}$ luôn có $(F^*_{R''})^+ \neq (F^*_{R'})^+$.

- Bước 0: Đặt $K_0 = R$, tính ${F_R^*}^+$.

- Bước i $(1 \leq i \leq q-1)$:

$$K_i = \begin{cases} K_{i-1} \setminus \{r_i\} & \text{nếu } \left(F^*_{K_{i-1}\setminus\{r_i\}}\right)^+ = {F_R^*}^+, \\ K_{i-1} & \text{nếu ngược lại.} \end{cases}$$

- Bước q:

Nếu K_{q-1} chỉ còn r_q, thì đặt $K_q = K_{q-1}$.

Nếu K_{q-1} chứa không chỉ có r_q, thì đặt

$$K_q = \begin{cases} K_{q-1} \setminus \{r_q\} & \text{nếu } \left(F^*_{K_{q-1}\setminus\{r_q\}}\right)^+ = {F_R^*}^+, \\ K_{q-1} & \text{nếu ngược lại.} \end{cases}$$

- Bước $q+1$: Đặt $R' = K_q$.

Định lý 4.2. *Thuật toán 4.1 là dừng và cho kết quả là tập luật R' không dư thừa.*

Chứng minh. Chúng ta sẽ chứng minh bằng phương pháp phản chứng.

Lưu ý rằng để có được K_{q-1} chúng ta đã kiểm tra tính dư thừa của $q-1$ luật là $r_1, \ldots, r_{q-1}$, do đó, như thuật toán đã chỉ rõ, có thể xảy ra các khả năng sau:

- Khả năng thứ nhất, K_{q-1} chỉ chứa một phần tử. Thế thì phần tử này chính là r_q và do đó K_{q-1} không thể "lùi" đi đâu được nữa. Vậy thì $K_q = K_{q-1}$, tức là $R' = K_q = \{r_q\}$.

- Khả năng thứ hai: K_{q-1} có ít nhất hai phần tử. Thế thì ngoài r_q ra, K_{q-1} còn chứa ít nhất một phần tử nữa. Khi đó, theo thuật toán chúng ta có:

$$K_q = \begin{cases} K_{q-1} \setminus \{r_q\} & \text{nếu } \left(F^*_{K_{q-1}\setminus\{r_q\}}\right)^+ = {F_R^*}^+, \\ K_{q-1} & \text{nếu ngược lại.} \end{cases}$$

Giả sử ngược lại rằng K_q chưa phải là tối ưu, tức là $R' \subset K_q$ và $R' \neq K_q$. Điều đó có nghĩa là trong K_q vẫn còn luật thừa, nói cách khác, $\exists r \in K_q$ sao cho với $R'' = K_q \setminus \{r\}$ thì $(F^*_{R''})^+ = (F^*_{K_q})^+$. Xét từng trường hợp đối với K_q:

(1) $K_q = K_{q-1} \setminus \{r_q\}$: thế thì mọi luật trong tập R đã được kiểm tra hết, điều này mâu thuẫn với việc trong K_q ngoài r_q ra vẫn còn ít nhất một luật nào đó chưa kiểm tra.

(2) $K_q = K_{q-1}$: trong trường hợp này, theo thuật toán thì $\left(F^*_{K_{q-1}\setminus\{r_q\}}\right)^+ \neq {F_R^*}^+$, tức là r_q không phải là luật thừa và như vậy tất cả các luật thuộc R đã được kiểm tra. Điều này lại mâu thuẫn với việc trong K_q vẫn còn luật thừa.

Như vậy điều giả thiết rằng $R' \subset K_q$ là sai. Thuật toán được chứng minh.

Trên cơ sở thuật toán tính bao đóng (Thuật toán 3.1), có thể chứng minh được kết quả sau.

Mệnh đề 4.3. *Thuật toán sàng lọc sự dư thừa của tập luật nêu trên có độ phức tạp là đa thức theo lực lượng của F và R.*

Nhận xét. Nếu thay đổi thứ tự của các luật trong dãy $R = \langle r_1, \dots, r_q \rangle$, thì Thuật toán 4.1 sẽ cho một hệ luật không dư thừa khác.

Dễ dàng thấy rằng, để kiểm tra tính dư thừa của một hệ luật, chúng ta có thuật toán sau.

Thuật toán 4.4. (sàng lọc sự dư thừa trong hệ luật)

Cho hệ luật $L = \langle F, R \rangle$ với $F = \langle f_1, \dots, f_p \rangle$, $R = \langle r_1, \dots, r_q \rangle$ và F^* là tập các sự kiện chỉ tham gia trong vế trái mà không tham gia trong vế phải của các luật. Khi đó, để sàng lọc dư thừa của hệ luật L, chúng ta sẽ tiến hành các bước sau:

- Dùng Thuật toán 4.1 loại các luật không cần thiết: từ $L = \langle F, R \rangle$ có $L' = \langle F, R' \rangle$, trong đó R' là tập luật không dư thừa.
- Xây dựng hệ luật không dư thừa $L'' = \langle F', R' \rangle$, với $F' = F \setminus (F^*_{R'})^+$.

5. NHỮNG VẤN ĐỀ MỞ

Như vậy, chúng ta đã xây dựng thuật toán tìm bao đóng của tập sự kiện và loại bỏ dư thừa của tập luật trong hệ luật dạng 1. Các thuật toán đó có ý nghĩa và đóng vai trò quan trọng trong quá trình suy diễn dựa vào hệ luật của hệ chuyên gia.

Dưới đây đề cập một số vấn đề có thể quan tâm nghiên cứu.

Vấn đề 1. Tìm thuật toán tính bao đóng và sàng lọc sự dư thừa đối với các hệ luật dạng 2.

Vấn đề 2. Xây dựng thuật toán tìm tập luật tối ưu (đối với cả hai dạng luật), theo nghĩa sau đây. Giả sử có hệ luật $L = \langle F, R \rangle$ với $F = \{f_1, \dots, f_p\}, R = \{r_1, \dots, r_q\}$ và F^* là tập các sự kiện chỉ tham gia trong vế trái mà không tham gia trong vế phải của các luật. Hãy tìm cách xác định tập G các luật tương thích với F và F^* sao cho các điều sau thỏa mãn:

(i) $F_G^{*+} = F_R^{*+}$,

(ii) với mọi tập luật I mà $F_I^{*+} = F_R^{*+}$ thì $|G| \leq |I|$, ở đây $|I|$ là ký hiệu số phần tử của tập I.

Lời cảm ơn. Tác giả xin chân thành cảm ơn PGS TSKH Nguyễn Xuân Huy và PGS TS Vũ Đức Thi đã đóng góp những ý kiến quí báu trong quá trình hoàn thành bài báo này. Tác giả cũng xin cảm ơn TS Ngô Quốc Tạo đã đọc và góp ý kiến cho bản thảo bài báo.

TÀI LIỆU THAM KHẢO

[1] Bạch Hưng Khang, Hoàng Kiếm, *Trí tuệ nhân tạo: các phương pháp và ứng dụng*, NXB Khoa học và Kỹ thuật, 1989.

[2] Durkin J., *Expert Systems*, Prentice Hall, 1994.

[3] Maier D., *The Theory of Relational Databases*, Computer Science Press, 1983.

[4] Sundermeyer K., *Knowledge Based Systems*, Wissenschafts Verlag, 1991.

[5] Turban E., *Decisions Support and Expert Systems - Management Support Systems*, Prentice Hall, 1998.

Nhận bài ngày 25 - 8 - 2000

Viện Công nghệ thông tin