ẠP CHÍ KHOA HỌC No 3 - 1991

ặng Huy Ruận

HỮNG DÃY BIỂU THỨC HÍNH QUY SUY RỘNG Ó ĐỘ PHỨC TẠP OTOMAT CAO

Biểu thức chính quy suy rộng trong bảng chữ cái A là biểu thức được xây dựng từ các biểu ức cơ bản (ϕ, $\wedge$ và $a \in A$) nhờ các phép nhân ghép (.), lặp (*), hợp ($\cup$), giao ($\cap$) và lấy phần [1].

Số các vị trí của những ký hiệu thuộc bảng chữ cái A xuất hiện trong biểu thức B được gọi độ dài của biểu thức này và ký hiệu bằng $|B|$.

Số trạng thái tối thiểu đủ để xây dựng ôtômát tiền định đoán nhận tập từ được cho bởi biểu ức B được gọi là độ phức tạp ôtômát của biểu thức này và ký hiệu bằng $G(B)$.

Đối với biểu thức chính quy suy rộng tùy ý B có độ sâu đặt dấu phần bù t, độ dài chính xác B) thỏa mãn quan hệ:

$$\underbrace{\log_2 \log_2 \ldots \log_2}_{t+1 \text{ lần}} G(B) \leq \beta(B)$$

Mặt khác có thể xây dựng những dãy biểu thức chính quy suy rộng, mà ôtômát tiền định án nhận những tập từ được cho bởi chúng đòi hỏi một số trạng thái đủ lớn. Thật vậy, có kết ả sau:

Định lý. Đối với các số tự nhiên tùy ý s, t có thể xây dựng được dãy các biểu thức chính y suy rộng $B_{s,t,n}$ trong bảng chữ cái gồm ba ký hiệu, sao cho

1) Với n tùy ý biểu thức $B_{s,t,n}$ chứa t dấu phần bù và có độ dài không vượt quá n.

2) Với bất kỳ hằng số $C > 2^s$ nào khi n đủ lớn

$$\underbrace{\log_b \log_b \ldots \log_b}_{t+1 \text{ lần}} G(B_{s,t,n}) \geq \frac{n}{C \log_2 n}, \tag{1}$$

ong đó

$$b = \sqrt{\frac{(2^s)^{2^s}}{(2^s-1)^{2^s-1}}}.$$

Chứng minh định lý này gồm một số bước. Trước hết, đối với số t tùy ý xây dựng biểu thức ính quy suy rộng trong bảng chữ cái gồm t+5 ký hiệu. Sau đó thu hẹp số ký hiệu của bảng chữ i xuống còn 3. Cuối cùng, tính độ dài của các biểu thức đã được xây dựng và chỉ ra rằng, với đủ lớn số các phần dư của những biểu thức này và độ dài của chúng thỏa mãn quan hệ (1).

1. XÂY DỰNG BIỂU THỨC CHÍNH QUY SUY RỘNG TRONG BẢNG CHỮ CÁI GỒM T+5 KÝ HIỆU

1. Giả sử t là số tự nhiên nào đó. Ký hiệu t+1 bằng k và $L_1 = \{0, 1, x, \beta_1, \beta_2, \dots, \beta_k\}$ l bảng chữ cái nào đó.

Chọn các số tự nhiên m_0, s và tập hợp D_0 gồm $2^s m_0$ từ thuộc bảng chữ cái $\{0, 1\}$. Xác địn các tập từ $D_1, D_2, \dots, D_k$ bằng quy nạp.

Giả sử D_i đã được xác định và gồm $2^s m_i$ phần tử. Khi đó D_{i+1} là tập gồm tất cả các tậ con của D_i, mà mỗi tập con này có chứa đúng m_i phần tử. Tập D_{i+1} gồm $C^{m_i}_{2^s m_i} = 2^s C^{m_i-1}_{2^s m_i-}$ phần tử. Đặt $m_{i+1} = \frac{1}{2^s} C^{m_i}_{2^s m_i} = C^{m_i-1}_{2^s m_i - 1}$. Dùng ℓ^i và h^i với chỉ số dưới hoặc không để ký hiệ phần tử thuộc D_i.

2. Xây dựng các biểu thức chính quy suy rộng

Trước hết xây dựng các biểu thức phụ:

$$\begin{aligned} \mathcal{R}_i &= (0 \cup 1 \cup \beta_0 \cup \beta_1 \cup \cdots \cup \beta_i)^*, \\ \mathcal{E}_i &= \beta_i \cup \beta_{i+1} \cup \cdots \cup \beta_k, \quad (0 \le i \le k) \\ L_1^* &= (0 \cup 1 \cup x \cup \mathcal{E}_0)^*. \end{aligned}$$

Các biểu thức chính quy suy rộng được xây dựng bằng quy nạp như sau:

$$\begin{aligned} B_0 &= \bigcup_{\ell^0 \in D_0} \ell^0 \mathcal{E}_0 L_1^* \mathcal{E}_0 \ell^0, \\ B_{i+1} &= C(\mathcal{R}_i \beta_i B_i \beta_i \mathcal{R}_i) \quad (0 \le i \le k) \end{aligned}$$

3. Các từ đặc trưng

Đối với mỗi $\ell^i \in D_i$ xây dựng các từ $\lambda_i(\ell^i)$, $\overline{\lambda}_i(\ell^i)$ bằng quy nạp theo i:

a) $\lambda_0(\ell^0) = \overline{\lambda}_0(\ell^0) = \ell^0$,

b) Giả sử ℓ^{i+1} là phần tử nào đó thuộc D^{i+1}. Đối với mỗi $\ell^i \in \ell^{i+1}$ đã xây dựng được t $\beta_i\ \lambda_i(\ell^i)$ $(\overline{\lambda}_i(h^i)\beta_i)$. Rải tất cả các từ $\beta_i \lambda_i(\ell^i)$ $(\overline{\lambda}_i(h^i)\beta_i)$ theo một thứ tự nào đó (chẳng hạn, th tự tự điển), rồi lấy tích ghép của tất cả các từ này. Từ nhận được ký bằng $\lambda_{i+1}(\ell^{i+1})$ $(\overline{\lambda}_{i+1}(\ell^{i+1})$

4. Các tập phần dư và các tính chất của chúng

Đối với phần tử $\ell^i \in D_i$ xây dựng tập $K_i(\ell^i)$ quy nạp theo i bằng cách sau:

$$\begin{aligned} K_0(\ell^0) &= L_1^* \mathcal{E}_0 \ell^0, \\ K_{i+1}(\ell^{i+1} &= C(\bigcup_{\ell^i \in \ell^{i+1}} K_i(\ell^i)) \beta_i \mathcal{R}_i \quad (0 \le i < k). \end{aligned}$$

Bổ đề 1. Đối với mọi từ X và các số bất kỳ ℓ^i, h^i $(0 \le i < k)$, $s (s \ge 1)$

$$X \beta_s \overline{\lambda}_i(h^i) \in K_i(\ell^i) \equiv h^i = \ell^i.$$

Từ bổ đề trên ta có hệ quả.

Hệ quả 1. Tồn tại $2^{*}.m_k$ tập khác nhau dạng

$$\mathcal{K}_k(\ell^k).$$

Thực hiện phép chia bên trái tập $\mathcal{B}_i$ $(0 \leq i \leq k)$ cho từ dạng đặc biệt ta có

Bổ đề 2. Đối với từ tùy ý Y và số bất kỳ ℓ^i $(0 \leq i \leq k)$ nếu

$$Y \in \lambda_i(\ell^i)\mathcal{E}_i L_1^* \quad \text{thì} \quad \frac{\mathcal{B}_i}{Yx} = \mathcal{K}_i(\ell^i).$$

Từ bổ đề 2 suy ra: với số ℓ^k tùy ý thuộc D_k tập $\mathcal{B}_k$ có tập phần dư dạng $\mathcal{K}_k(\ell^k)$, còn số các tập dư dạng $\mathcal{K}(\ell^k)$ theo hệ quả 1, bằng $2_* m_k$. Mặt khác số trạng thái tối thiểu của ôtômát đoán nhận $\mathcal{B}_k$ lại bằng đúng số tập phần dư của nó.

Khi có ôtômát đoán nhận tập từ nào đó, muốn tìm ôtômát đoán nhận phần bù của nó ta chỉ việc đổi, trong ôtômát đã biết, các trạng thái kết thành không kết và ngược lại.

Đặt $\qquad \mathcal{A}_t = \mathcal{R}_t \beta_t \mathcal{B}_t \beta_t \mathcal{R}_t.$

Do $k = t + 1$, nên $\mathcal{B}_k = C\mathcal{A}_t$, ta có hệ quả

Hệ quả 2. $G(\mathcal{A}_t) = G(\mathcal{B}_k) \geq 2^{*}.m_k = 2^{*}_{m_{t+1}}$.

2. XÂY DỰNG BIỂU THỨC CHÍNH QUY SUY RỘNG TRÊN BẢNG CHỮ CÁI GỒM BA KÝ HIỆU

Xây dựng bảng chữ cái $L_2 = \{0, 1, \alpha\}$ và ánh xạ của tập các từ thuộc L_1 vào tập từ trong L_2:

$$\varphi(0) = 0, \quad \varphi(1) = 1; \quad \varphi(x) = \alpha_0\alpha; \quad \varphi(\beta_i) = \alpha O^{i+2}\alpha \quad (0 \leq i \leq k)$$

Bổ đề 3. Đối với các tập từ $\mathcal{M} \subseteq L_2^*$ và $\mathcal{N} \subseteq L_1^*$, nếu $\mathcal{M} \cap \varphi(L_1^*) = \varphi(\mathcal{N})$, thì

$$G(\mathcal{M}) \geq G(\mathcal{N}),$$

Các biểu thức chính quy phụ:

$$\overline{\mathcal{E}}_0 = \bigcup_{i=0}^{k} \alpha O^{i+3}\alpha, \quad \overline{\mathcal{L}}_0 = (0 \cup 1 \cup \alpha_0\alpha \cup \overline{\mathcal{E}}_0)^*,$$

$$\overline{\mathcal{R}}_i = (O \cup 1 \cup \bigcup_{r=0}^{i} \alpha O^{r+3}\alpha)^* \quad (0 \leq i \leq k)$$

Các biểu thức chính quy suy rộng cần tìm được xây dựng theo phương pháp quy nạp như sau:

$$\overline{\mathcal{B}}_0 = \bigcup_{\ell^0 \in D_0} \ell^0 \overline{\mathcal{E}}_0 \overline{\mathcal{L}}_0 \overline{\mathcal{E}}_0 \ell^0,$$

$$\overline{\mathcal{B}}_{i+1} = C(\overline{\mathcal{R}}_i \alpha O^{i+3}\alpha \overline{\mathcal{B}}_i \alpha O^{i+3}\alpha \overline{\mathcal{R}}_i) \quad (0 \leq i < k),$$

$$\overline{\mathcal{A}}_t = \overline{\mathcal{R}}_t \alpha O^{t+3}\alpha \overline{\mathcal{B}}_t \alpha O^{t+3}\alpha \overline{\mathcal{R}}_t.$$

Vì $k = t + 1$, nên ta có $\overline{B}_k = C\overline{A}_t$ và

$$G(\overline{B}_k) = G(\overline{A}_t). \tag{2}$$

Bổ đề 4. Đối với chỉ số i tùy ý $(0 \leq i \leq k)$

$$\overline{B}_i \cap \varphi(L_1^*) = \varphi(B_i).$$

Từ các bổ đề 3, 4, hệ quả 2 và quan hệ (2) ta có:

Hệ quả 3. $\quad G(\overline{A}_t) \geq 2^s m_{t+1}$.

Sau khi chọn D_0 một cách thích hợp theo m_0 và tính độ dài biểu thức $\overline{A}_t$ được ước lượng

$$|\overline{A}_t| \geq 2^{s+1} m_0 \, (\log_2 m_0 + C(t)),$$

Đối với số tự nhiên n nào đó đã cho, ta chọn số tự nhiên m_0 lớn nhất sao cho

$$2^{s+1} m_0 \, (\log_2 m_0 + C(t)) \geq n.$$

Khi đó $|\overline{A}_t \leq n$.

Do n tăng đại lượng m_0 dần tới vô hạn, nên đối với hằng số tùy ý $C > 2^s$ khi n đủ lớn

$$\underbrace{\log_b \log_b \ldots \log_b}_{t+1 \text{ lần}} G(\overline{A}_t) > \frac{n}{C\log_2 n}$$

Nếu đặt $B_{s,t,n} = \overline{A}_t$ ta được quan hệ (1).

TÀI LIỆU THAM KHẢO

1. Данг Гуи Руан, ДАН СССР, No 1, T. 213, (1973).

Dang Huy Ruan

ON GENERALIZED REGULAR EXPRESSION HAVING HIGH AUTOMATION COMPLEXITY

The paper presents the way of building series of regular expressions. To accept languages defined by there regular expressions. It is necessary to have automata having larger number of states.

Khoa Toán Cơ Tin học - ĐHTH Hà Nội