ĐỘ PHỨC TẠP Ô-TÔ-MÁT CỦA CÁC DÃY BIỂU THỨC CHÍNH QUY SUY RỘNG

Đặng Huy Ruận
Khoa Toán - Cơ - Tin học, ĐHTH Hà Hội

Giả sử có bảng chữ cái A. Dãy vô hạn các ký hiệu thuộc A được gọi là một siêu từ trên bảng chữ cái A. Tập hợp tất cả các siêu từ trên bảng chữ cái A ký hiệu bằng A^∞ [1].

Tập con tùy ý của tập A^∞ được gọi là một siêu ngôn ngữ trên bảng chữ cái A.

Tích của ngôn ngữ M_1 với siêu ngôn ngữ M_2 trên bảng chữ cái A là một siêu ngôn ngữ, gồm tất cả các siêu từ dạng $\alpha = a(1)a(2)\ldots$ sao cho có số tự nhiên i nào đó, để từ $a(1)$, $a(2)\ldots$ $a(i)$ thuộc M_1, còn siêu từ $a(i+1)$ $a(i+2)$... thuộc M_2.

Siêu lặp của ngôn ngữ trên bảng chữ cái A (ký hiệu bằng M^∞) là siêu ngôn ngữ trên bảng chữ cái A, gồm tất cả các siêu từ $a(1)$ $a(2)$..., sao cho đối với dãy số tự nhiên tăng nào đó $i_1, i_2, i_3, \ldots$ thỏa mãn quan hệ.

$$a(i_t)a(i_{t+1})\ldots a(i_{t+1}-1) \in M \quad t = 1, 2, \ldots$$

Quy ước rằng $\phi^\infty = \phi$ và đối với các siêu ngôn ngữ tùy ý M đều có $\phi.M = \phi$.

Biểu thức chính quy suy rộng (B.c.s.) trên bảng chữ cái A là biểu thức tùy ý được xây dựng từ các biểu thức cơ bản (ϕ, Λ và $a \in A$) nhờ các phép toán nhân (.), lặp (*) và các phép toán tập hợp: Hợp ($\cup$), giao ($\cap$) và lấy phần bù (C).

Lớp các biểu thức chính quy suy rộng siêu ngôn ngữ (B.c.s.s.n.) trên bảng chữ cái A được xác định như sau:

1) Nếu R - B.c.s., xác định siêu ngôn ngữ nào đó, thì R là B.c.s.s.n. cơ bản;

2) Nếu R - B.c.s., thì R^∞ là R.c.s.s.n.;

3) Nếu R_1 và R_2 là B.c.s.s.n., thì $R_1 \cup R_2$ là B.c.s.s.n.;

4) Nếu R_1 - B.c.s., còn R_2 - B.c.s.s.n.; thì R_1, R_2 là B.c.s.s.n.;

5) Nếu R - B.c.s.s.n.; Thì CR là B.c.s.s.n.;

6) Nếu R_1, R_2 là B.c.s.s.n., thì $R_1 \cap R_2$ là B.c.s.s.n.;

7) Chỉ các biểu thức định nghĩa theo các mục 1-6 mới là B.c.s.s.n.

Giả sử $\mathcal{M}$ là B.c.s.s.n. trên bảng chữ cái A. Số các vị trí của các ký hiệu thuộc A, chứa trong $\mathcal{M}$ được gọi là độ dài của biểu thức $\mathcal{M}$ và ký hiệu bằng $|\mathcal{M}|$.

Số trạng thái ít nhất đủ để xây dựng ô-tô-mát đơn định đoán nhận siêu ngôn ngữ được cho bởi biểu thức $\mathcal{M}$ được gọi là độ phức tạp ô-tô-mát của biểu thức $\mathcal{M}$ và được ký hiệu bằng $G(\mathcal{M})$.

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

1) Với n tùy y biểu thức $C_{s,t,n}$ 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

$$\log_b \log_b \ldots \log_b G(\mathcal{C}_{s,t,n}) \geq \frac{n}{C \log_2 n} \tag{1}$$

trong đó:

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

Chứng minh định lý trên gồm một số bước. Trước hết, đối với số t tùy ý xây dựng B.c.s.s.n. trong bảng chữ cái gồm $t+6$ ký hiệu. Sau đó thu hẹp số ký hiệu xuống còn 3. Cuối cùng, tính độ dài của các biểu thức đã được xây dựng và chỉ ra rằng, với n đủ lớn số các phần dư của những biểu thức này và độ dài của chúng thỏa mãn bất đẳng thức (1).

§1. XÂY DỰNG BIỂU THỨC CHÍNH QUY SUY RỘNG TRÊN BẢNG CHỮ CÁI GỒM $t+5$ KÝ HIỆU

1. Giả sử t là số tự nhiên nào đó. Ký hiệu $t+\ell$ bằng k và $L_1 = \{0, 1, x, \beta_0, \beta_1, \ldots, \beta_k\}$ là bảnng chữ cái nào đó.

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

Giả sử D_i đã được xác định gồm $2^s.m_i$ phần tử. Khi đó D_{i+1} là tập gồm tất cả các tập con của D_i, mà mỗi tập con này có chứa đúng m_i phần tử. Tập D_{i+1} gồm

$$C^{m_i}_{2^s.m_i} = 2^s.C^{m_i - 1}_{2^s.m_i - 1} \quad \text{phần tử. Đặt} \quad m_{i+1} = \frac{1}{2^s} C^{m_i}_{2^s.m_i} = C^{m_i - 1}_{2^s.m_i - 1}$$

Dùng ℓ^i và h^i với chỉ số dưới hoặc không để ký hiệu phần tử thuộc D_i.

2. Xây dựng các biểu thức chính quy suy rộng trên L_1

Trước hết xây dựng các biểu thức phụ:

$$\mathcal{R}_i = (0 \cup 1 \cup \beta_0 \cup \beta_1 \cup \cdots \cup \beta_i)^*$$

$$\mathcal{E}_i = \beta_i \cup \beta_{i+1} \cup \cdots \cup \beta_k, \quad (0 \leq i \leq k)$$

$$L_1^* = (0 \cup 1 \cup x \cup \mathcal{E}_0)^*$$

Các B.c.s. được xây dựng bằng quy nạp như sau:

$$B_0 = \bigcup_{\ell^0 \in D_0} \ell^0 \mathcal{E}_0 L_1^* \mathcal{E}_0 \ell^0,$$

$$B_{i+1} = C(\mathcal{R}_i \beta_i B_i \beta_i \mathcal{R}_i) \quad (0 \leq i < k)$$

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 tùy ý thuộc D_{i+1} và đối với mỗi $\ell^i \in \ell^{i+1}$ và $h^i \in \ell^{i+1}$ đã xây dựng được từ $\beta_i\lambda_i(\ell^i)$ $(\overline{\lambda}_i(h^i)\beta_i)$. 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, thứ tự tự điển), rồi lấy tích ghép của tất cả các từ này. Từ nhận được ký hiệu bằng

$$\lambda_{i+1}(\ell^{i+1})\ (\overline{\lambda}_{i+1}(\ell^{i+1})).$$

4. Các tập phần dư và các tính chất

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

$$K_0(\ell^0) = L_1^* \mathcal{E}_0 \ell^0,$$

$$K_{i+1}(\ell^{i+1}) = C\Big(\bigcup_{\ell^i \in \ell^{i+1}} K_i(\ell^i)\Big)\beta_i \mathcal{R}_i \quad (0 \le i < k).$$

Bổ đề 1. Đối với mỗi từ $X \in L_1^*$ và các số bất kỳ ℓ^i, h^i $(0 \le i \le k)$, s $(s \ge i)$

$$X\beta_s\overline{\lambda}_i(h^i) \in K_i 9\ell^i) \equiv h^i = \ell^i.$$

§2. XÂY DỰNG BIỂU THỨC CHÍNH QUY SUY RỘNG SIÊU NGÔN NGỮ TRÊN BẢNG CHỮ CÁI GỒM $t+6$ KÝ HIỆU VÀ CÁC TÍNH CHẤT CỦA CHÚNG

Xây dựng bảng chữ cái $L_2 = L_1 \cup \{y\}$, trong đó y $y \overline{\in} L_1$.

Đối với mỗi i $(0 \le i \le k)$ B.c.s. $\mathcal{B} - i$ là tập phần dư $K_i(\ell^i)$ đã được xây dựng trên bảng chữ cái L_1, biểu thức chính quy suy rộng siêu ngôn ngữ C_i và tập phần dư $\mathcal{H}_i(\ell^i)$ trên bảng chữ cái L_2 có dạng:

$$\begin{aligned} C_i &= \mathcal{B}_i\{y\}^\infty \\ \mathcal{H}_i(\ell^i) &= K_i(\ell^i).\{y\}^\infty \end{aligned} \qquad (0 \le i \le k)$$

Bổ đề 2. Đối với mỗi từ $X \in L_1^*$, siêu từ tùy ý $Y \in \{y\}^\infty$ và các số bất kỳ ℓ^i, h^i $(0 \le i \le k)$, s $(s \ge i)$

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

Từ các bổ đề 1, 2 suy ra:

Hệ quả 1. Tồn tại $2^s m_k$ tập khác nhau dạng $\mathcal{H}_i(\ell^k)$.

Thực hiện phép chia bên trái các siêu ngôn ngữ C_i $(0 \le i \le k)$ cho từ đặc biệt ta có:

Bổ đề 3. Đối với từ tùy ý Z và số bất kỳ ℓ^i $(0 \le i \le k)$, nếu $Z \in \lambda_i(\ell^i)\mathcal{E}_i L_1^*$, thì

$$C_i/Zx = \mathcal{H}_i(\ell^i)..$$

Do bổ đề 3 suy ra: Với số ℓ^k tùy ý thuộc D_k tập C_k có tập phần dư dạng $\mathcal{H}_k(\ell^k)$. Mặt khác số các tập siêu từ dạng $\mathcal{H}_k(\ell^k)$, theo hệ quả 1, bằng $2^s.m_k$, còn số trạng thái để xây dựng ô-tô-mat đơn định đoán nhận C_k không thể ít hơn số phần dư của nó. Ta có hệ qua sau:

Hệ quả 2.

$$G(C_k) \ge 2^s.m_k$$

§3. XÂY DỰNG BIỂU THỨC CHÍNH QUY SUY RỘNG SIÊU NGÔN NGỮ TRÊN BẢNG CHỮ CÁI GỒM BA KÝ HIỆU

1. **Xây dựng bảng chữ cái $L_3 = \{0, 1, \alpha\}$ và ánh xạ φ của tập gồm các từ thuộc L_2 vào tập từ trong L_3:**

$$\varphi_0 = 0; \quad \varphi_1 = 1; \quad \varphi(x) = \alpha 0 \alpha; \quad (y) = \alpha 0^2 \alpha; \quad \varphi(\beta_i) = \alpha \ell^{i+1} \alpha \quad (0 \le i \le k).$$

Bổ đề 4. Đối với các tập tùy ý $\mathcal{M} \subseteq L_3^\infty$ và $\mathcal{N} \subseteq L_2^\infty$, nếu $\mathcal{M} \cap \varphi(L_2^\infty) = \varphi(\mathcal{N})$, thì $G(\mathcal{M}) \ge G(\mathcal{N})$.

2. Xây dựng biểu thức chính quy suy rộng siêu ngôn ngữ

Các biểu thức chính quy:

$$\overline{\mathcal{E}}_0 = \bigcup_{i=0}^{k} \alpha 0^{i+3} \alpha,$$

$$\overline{\mathcal{L}}_0 = (0 \cup 1 \cup \alpha 0 \alpha \cup \overline{\mathcal{E}}_0)^*,$$

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

Các B.c.s. và b.c.s.s.n. được xây dựng bằng quy nạp như sau:

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

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

$$\overline{\mathcal{C}}_i = \overline{\mathcal{B}}_i . \{\alpha 0^2 \alpha\}^\infty \quad (0 \le i \le k)$$

Bổ đề 5. Đối với mỗi số i tùy ý $(0 \le i \le k)$

$$\overline{\mathcal{C}}_i \cap \varphi(L_2^\infty) = \varphi(\mathcal{C}_i)$$

Từ hệ quả 2 các bổ đề 4, 5 ta có:

Hệ quả 3.

$$G(\overline{\mathcal{C}}_k \ge 2^s m_k.$$

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

$$|\overline{\mathcal{C}}_k| \le 2^{s+1} m_0 \, (\log_2 m_0 + c(k)).$$

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

$$2^{s+1} m_0 (\log_2 m_0 + c(k)) \le n.$$

Khi đó $|\overline{\mathcal{C}}_k| \le n$.

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.

$$\log_b \log_b \ldots \log_b G(\overline{\mathcal{C}}_k) > \frac{n}{C \log_2 n}.$$

TÀI LIỆU THAM KHẢO

1. Кудрявцев В. Б., Алёшин С. Б., Подколзин А. С. Введение в теорию автоматов. Москва "Наука" 1985.

VNU,H Journal of science, Nat. sci. t.XI, n°1, 1995

ON THE COMPLEXITY OF A FINITE AUTOMATON CORRESPONDING TO SEQUENCES OF GENERALIZED REGULAR EXPRESSIONS

Dang Huy Ruan
Faculty of Mathematics, Hanoi University

Suppose A is a finite alphabet. An infinite sequence of symbols in the alphabet A is called a meta-word.

The set of all meta-words in the alphabet A is denoted by A^{∞}.

A generalized regular expression in the alphabet A is an expression constructed from elementary expressions (ϕ, Λ and $a \in A$) by means of the operations of multiplication (.), iteration ($*$), meta-iteration (∞), set-theoretic union ($\cup$), intersection ($\cap$) and complementation (C).

The smallest number of states of a deterministic automaton which recognizes meta-language generated by a expression M is called the complexity of a finite automaton corresponding to the expression M.

The paper has constructed comsequences of generalised regular expressions which have the complexity of a finite automaton is rather large.