Tạp chí Tin học và Điều khiển học, T.22, S.1 (2006), 1–13

HỌ TRỪU TƯỢNG CÁC NGUỒN

NGUYỄN THẾ BÌNH

Bộ môn Tin học, Trường Đại học Dân lập Hải Phòng

Abstract. This article presents the basic concepts of abstract family of sources and proves that family of languages defined by abstract family of sources coincides with family of languages accepted by abstract family of automata.

Tóm tắt. Bài báo trình bày những khái niệm cơ bản về họ trừu tượng các nguồn và chứng minh rằng họ các ngôn ngữ được xác định bởi họ trừu tượng các nguồn trùng với họ các ngôn ngữ được đoán nhận bởi họ trừu tượng các otomat.

1. ĐẶT VẤN ĐỀ

Ngôn ngữ hình thức và otomat là cơ sở quan trọng của tin học đang phát triển mạnh mẽ theo nhiều hướng khác nhau ([1, 2, 6, 7, 8]). Một trong những hướng phát triển đó là xây dựng mô hình nguồn của otomat hữu hạn, các phép toán và các tính chất của nguồn ([1, 2]).

Bài báo trình bày các khái niệm cơ bản về họ trừu tượng các nguồn, đưa ra sự phân lớp họ trừu tượng các nguồn và chứng minh rằng họ các ngôn ngữ được xác định bởi họ trừu tượng các nguồn trùng với họ các ngôn ngữ được đoán nhận bởi họ trừu tượng các otomat.

2. CÁC KHÁI NIỆM CƠ BẢN

2.1. Bảng chữ cái, xâu, ngôn ngữ ([2, 3])

Định nghĩa 1. Bảng chữ cái là tập các phần tử, mà mỗi phần tử được gọi là một ký hiệu.

Định nghĩa 2. Xâu (hay từ) trên bảng chữ cái là dãy các ký hiệu của bảng chữ cái đó được viết liền nhau.

Ghép tiếp hai xâu u và v, ký hiệu bởi uv, là một xâu được tạo ra bằng cách viết u xong viết v tiếp theo sau sao cho giữa chúng không có khoảng cách.

Độ dài của xâu ω là số các ký hiệu có trên xâu đó và được ký hiệu bởi $|\omega|$.

Xâu rỗng là xâu có độ dài bằng 0 và được ký hiệu bởi ε.

Định nghĩa 3. Ngôn ngữ trên bảng chữ cái là tập các xâu trên bảng chữ cái đó.

Tập tất cả các xâu trên bảng chữ cái Σ với độ dài khác không được ký hiệu bởi Σ^+.

Tập tất cả các xâu trên bảng chữ cái Σ với độ dài bất kỳ được ký hiệu bởi Σ^*.

Khi đó $\Sigma^* = \Sigma^+ \cup \{\varepsilon\}$.

Tập Φ và $\{\varepsilon\}$ được xem như là các ngôn ngữ trên bảng chữ cái bất kỳ.

2.2. Một số phép toán trên ngôn ngữ ([2, 3])

Giả sử L, L_1, L_2 là những ngôn ngữ bất kỳ trên bảng chữ cái Σ.

- Phép hợp:

Hợp của hai ngôn ngữ L_1 và L_2, được ký hiệu bởi $L_1 \cup L_2$, là tập:

$$L_1 \cup L_2 = \{\omega \in \Sigma^* \mid \omega \in L_1 \text{ hoặc } \omega \in L_2\}.$$

- Phép giao:

Giao của hai ngôn ngữ L_1 và L_2, được ký hiệu bởi $L_1 \cap L_2$, là tập:

$$L_1 \cap L_2 = \{\omega \in \Sigma^* \mid \omega \in L_1 \text{ và } \omega \in L_2\}.$$

- Phép ghép tiếp:

Ghép tiếp của hai ngôn ngữ L_1 và L_2, được ký hiệu L_1L_2, là tập:

$$L_1L_2 = \{uv \mid u \in L_1 \text{ và } v \in L_2\}.$$

- Phép lặp:

Lặp của ngôn ngữ L, được ký hiệu bởi L^*, là tập:

$$L^* = L^0 \cup L^1 \cup L^2 \cup \dots$$

trong đó L^i được xác định bởi công thức đệ quy như sau:

$$L^0 = \{\varepsilon\}, L^i = LL^{i-1} \text{ với mọi số nguyên } i > 0.$$

Lặp cắt của ngôn ngữ L, được ký hiệu bởi L^+, là tập:

$$L^+ = L^1 \cup L^2 \cup L^3 \cup \dots$$

- Phép đồng cấu:

Giả sử Σ_1, Σ_2 là các bảng chữ cái.

Ánh xạ h từ tập Σ_1^* vào tập Σ_2^* được gọi là đồng cấu nếu thỏa mãn điều kiện:

$$h(xy) = h(x)h(y) \quad \text{với} \quad x, y \in \Sigma_1^*.$$

Đối với ngôn ngữ L trên bảng chữ cái Σ_1 ta có:

$$h(L) = \{h(\omega) \mid \omega \in L\}.$$

Đồng cấu h được gọi là đồng cấu ε - tự do nếu $h(x) \neq \varepsilon$ với mọi $x \neq \varepsilon$.

- Phép đồng cấu ngược:

Giả sử h là một đồng cấu từ tập Σ_1^* vào tập Σ_2^*.

Đồng cấu ngược của h, ký hiệu bởi h^{-1}, là một ánh xạ từ tập Σ_2^* vào tập các tập con của Σ_1^* và được xác định như sau:

$$h^{-1}(y) = \{x \mid h(x) = y\}.$$

Đối với ngôn ngữ L trên bảng chữ cái Σ_2 ta có:

$$h^{-1}(L) = \{x \mid h(x) \in L\}.$$

2.3. Họ trừu tượng các ngôn ngữ AFL ([5])

Định nghĩa 4. Họ trừu tượng các ngôn ngữ là một bộ hai $(\Sigma, \mathcal{L})$ hay đơn giản là $\mathcal{L}$ khi Σ đã xác định và được xác định như sau:

1. Σ là bảng chữ cái, $\mathcal{L}$ là tập các ngôn ngữ trên Σ.

2. Đối với mỗi ngôn ngữ $L \in \mathcal{L}$, luôn tồn tại tập hữu hạn $\Delta \subset \Sigma$, sao cho $L \subseteq \Delta^*$.

3. $\mathcal{L}$ đóng đối với các phép hợp, ghép tiếp, lặp, đồng cấu ε- tự do, đồng cấu ngược và giao với các tập chính quy.

4. $L \neq \Phi$ với $L \in \mathcal{L}$.

Định nghĩa 5. Họ trừu tượng các ngôn ngữ $\mathcal{L}$ được gọi là đầy đủ nếu nó đóng với phép đồng cấu bất kỳ.

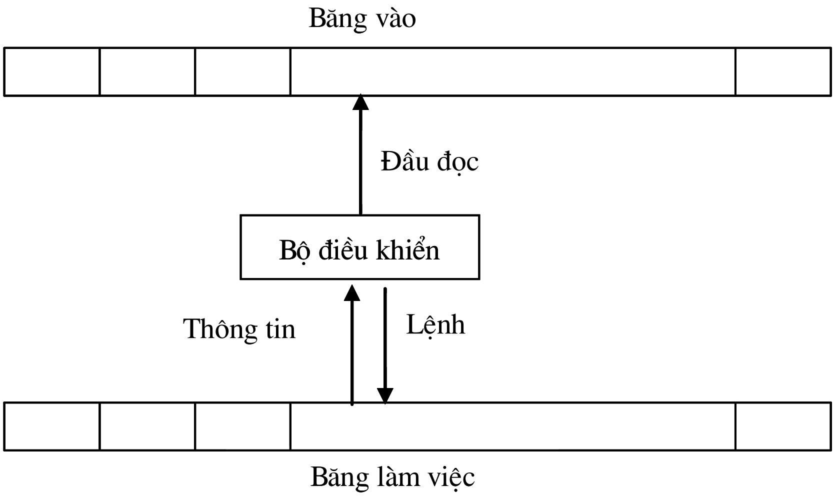
Ví dụ: Lớp các ngôn ngữ cảm ngữ cảnh là họ trừu tượng các ngôn ngữ. Còn lớp các ngôn ngữ chính quy, lớp các ngôn ngữ phi ngữ cảnh và lớp các ngôn ngữ ngữ cấu là họ trừu tượng đầy đủ các ngôn ngữ.

2.4. Họ trừu tượng các otomat AFA ([5])

Định nghĩa 6. Họ trừu tượng các otomat (một phía) là một bộ hai $(\Omega, \mathcal{A})$ hay đơn giản là $\mathcal{A}$ khi Ω đã xác định và được xác định như sau:
1. Ω là một bộ sáu $(K, \Sigma, \Gamma, I, f, g)$, trong đó:
1.1. K, Σ, Γ, I là các tập không trống,
1.2. f là ánh xạ từ tập $\Gamma^* \times I$ vào tập $\Gamma^* \cup \{\Phi\}$,
1.3. g là ánh xạ từ tập Γ^* vào tập các tập con hữu hạn của Γ^* và có tính chất sau:

$$g(\varepsilon) = \varepsilon \quad \text{và} \quad \varepsilon \in g(\gamma) \;\text{ khi và chỉ khi }\; \gamma = \varepsilon,$$

1.4. Đối với $\gamma \in g(\Gamma^*)(g(\Gamma^*) = \bigcup_{\alpha \in \Gamma^*} g(\alpha))$ luôn tồn tại $1_\gamma \in I$, sao cho $f(\gamma', 1_\gamma) = \gamma'$ với $\gamma \in g(\gamma')$, phần tử 1_γ thuộc I được gọi là lệnh đồng nhất (identical instruction),
1.5. Đối với $u \in I$ luôn tồn tại tập con hữu hạn $\Gamma_u \subseteq \Gamma$, sao cho nếu $\Gamma_1 \subseteq \Gamma$, $\gamma \in \Gamma_1^*$ và $f(\gamma, u) \neq \Phi$ thì $f(\gamma, u) \in (\Gamma_1 \cup \Gamma_u)^*$.
2. $\mathcal{A}$ là tập các otomat $A = (K_1, \Sigma_1, \delta, q_0, F)$, trong đó:
2.1. K_1 và Σ_1 là các tập con của K và Σ,
2.2. F là tập con của K_1 và $q_0 \in K_1$,
2.3. δ là ánh xạ từ tập $K_1 \times (\Sigma_1 \cup \{\varepsilon\}) \times g(\Gamma^*)$ vào tập các tập con hữu hạn $K_1 \times I$ với $G_A = \{\gamma \mid \delta(q, a, \gamma) \neq \Phi\}$ là tập hữu hạn.



Hình 1

Lưu ý: K là tập các trạng thái, Σ là tập các ký hiệu kết thúc, Γ là tập các ký hiệu không kết thúc, I là tập các lệnh, f là hàm biến đổi nội dung trên băng làm việc, g là hàm xác định xâu trên băng làm việc.

Otomat A có thể biểu diễn theo mô hình như Hình 1.

Bước chuyển của otomat A trên tập các hình trạng $K_1 \times \Sigma_1^* \times \Gamma^*$ (hoặc $K_1 \times \Gamma^*$) được xác định bởi quan hệ hai ngôi $\vdash$

$$(p, aw, \gamma) \vdash (p', w, \gamma') \text{ với } a \in \Sigma_1 \cup \{\varepsilon\}, (p', u) \in \delta(p, a, \gamma''),$$

$$\gamma'' \in g(\gamma) \text{ và } f(\gamma, u) = \gamma'.$$

Còn $\vdash^+, \vdash^*$ là bao đóng truyền ứng và bao đóng phản xạ truyền ứng của $\vdash$.

Định nghĩa 7. Ngôn ngữ được đoán nhận bởi otomat A là tập:

$$L(A) = \{w \in \Sigma_1^* \mid (q_0, w, \varepsilon) \vdash^* (q_f, \varepsilon, \varepsilon) \text{ với } q_f \in F\}.$$

Họ các ngôn ngữ được đoán nhận bởi họ trừu tượng các otomat là tập:

$$\mathcal{L}(\mathcal{A}) = \{L(A) \mid A \in \mathcal{A}\}.$$

Nhận xét 1

Từ Định nghĩa 6.1.5 ta có:

$$I_1 = \{u \mid (p', u) \in \delta(p, a, \gamma)\}, \Gamma_1 = \bigcup_{u \in I_1} \Gamma_u \text{ là các tập hữu hạn.}$$

Từ Định nghĩa 6.2.3 ta có:

$$G_A = \{\gamma \mid \delta(q, a, \gamma) \neq \Phi\} \text{ cũng là tập hữu hạn,}$$

nên δ là ánh xạ từ tập $K_1 \times (\Sigma_1 \cup \{\varepsilon\}) \times g(\Gamma^*)$ vào tập các tập con hữu hạn $K_1 \times I$ được xem là ánh xạ từ tập $K_1 \times (\Sigma_1 \cup \{\varepsilon\}) \times G_A$ vào tập các tập con hữu hạn $K_1 \times I_1$.

Giả sử T là hàm xác định trên tập tất cả các số nguyên không âm và nhận tất cả các giá trị là số nguyên dương.

$\mathcal{A}_{T(n)}$ là lớp các otomat $A = (K_1, \Sigma_1, \delta, q_0, F_1) \in \mathcal{A}$, sao cho với mọi $w \in \Sigma_1^*$ và $(p_0 w, \varepsilon) \vdash^* (p, \varepsilon, \gamma)$ ta luôn có

$$|\gamma| \leqslant T(|w|), T_{k(n)} = T(kn), \ \mathcal{A}_{Tk(n)} \text{ là lớp các otomat của } \mathcal{A} \text{ và } \mathcal{L}_{T(n)} = \bigcup_{k \geqslant 1} \mathcal{L}(\mathcal{A}_{Tk(n)}),$$

Trường hợp:

$$T(n) = n, \ \mathcal{L}(\mathcal{A}_{Tk(n)}) = \mathcal{L}(\mathcal{A}_{T(n)}) \text{ và } \mathcal{L}_{T(n)}(\mathcal{A}) = \mathcal{L}(\mathcal{A}_{T(n)}).$$

Khi đó $\bigcup_{k \geqslant 1} \mathcal{A}_{Tk(n)}$ là lớp các otomat tuyến tính giới nội và $\mathcal{L}_{T(n)}(\mathcal{A})$ là lớp các ngôn ngữ cảm ngữ cảnh.

Định lý 1. ([5]) *Nếu $\mathcal{A}$ là họ trừu tượng các otomat thì $\mathcal{L}(\mathcal{A})$ là họ trừu tượng đầy đủ các ngôn ngữ.*

Định lý 2. ([5]) *Nếu T là hàm tăng, thì $\mathcal{L}_{T(n)}(\mathcal{A})$ là họ trừu tượng các ngôn ngữ.*

Trên cơ sở định nghĩa họ trừu tượng các otomat và dựa vào đặc điểm của xâu trên băng làm việc, ta tách từ họ trừu tượng các otomat ra một số lớp:

Lớp các otomat loại 0

Otomat được gọi là otomat loại 0, nếu không có sự hạn chế và ràng buộc nào đối với xâu trên băng làm việc.

Otomat loại 0 được gọi là máy Turing.

Lớp các otomat loại 0, ký hiệu bởi $\mathcal{A}_0$, là tập các otomat loại 0.

Lớp các ngôn ngữ được đoán nhận bởi các otomat loại 0, ký hiệu bởi $\mathcal{L}(\mathcal{A}_0)$, là tập các ngôn ngữ được đoán nhận bởi các otomat loại 0.

Lớp các otomat loại 1

Otomat được gọi là otomat loại 1, nếu độ dài của xâu trên băng làm việc luôn bị giới hạn bởi hàm tuyến tính với biến là độ dài của xâu vào.

Otomat loại 1 được gọi là otomat tuyến tính giới nội.

Lớp các otomat loại 1, ký hiệu bởi $\mathcal{A}_1$, là tập các otomat loại 1.

Lớp các ngôn ngữ được đoán nhận bởi các otomat loại 1, ký hiệu bởi $\mathcal{L}(\mathcal{A}_1)$, là tập các ngôn ngữ được đoán nhận bởi các otomat loại 1.

Lớp các otomat loại 2

Otomat được gọi là otomat loại 2, nếu quá trình biến đổi nội dung của xâu trên băng làm việc luôn được thực hiện ở phần bên phải nhất (hoặc bên trái nhất).

Otomat loại 2 được gọi là otomat đẩy xuống.

Lớp các otomat loại 2, ký hiệu bởi $\mathcal{A}_2$, là tập các otomat loại 2.

Lớp các ngôn ngữ được đoán nhận bởi các otomat loại 2, ký hiệu bởi $\mathcal{L}(\mathcal{A}_2)$, là tập các ngôn ngữ được đoán nhận bởi các otomat loại 2.

Lớp các otomat loại 3

Otomat được gọi là otomat loại 3, nếu xâu trên băng làm việc luôn là xâu rỗng.

Otomat loại 3 được gọi là otomat hữu hạn.

Lớp các otomat loại 3, ký hiệu bởi $\mathcal{A}_3$, là tập các otomat loại 3.

Lớp các ngôn ngữ được đoán nhận bởi các otomat loại 3, ký hiệu bởi $\mathcal{L}(\mathcal{A}_3)$, là tập các ngôn ngữ được đoán nhận bởi các otomat loại 3.

Lớp các ngôn ngữ được đoán nhận bởi các otomat tương ứng với sự phân lớp trên thỏa mãn bao hàm thức sau:

$$\mathcal{L}(\mathcal{A}_3) \subset \mathcal{L}(\mathcal{A}_2) \subset \mathcal{L}(\mathcal{A}_1) \subset \mathcal{L}(\mathcal{A}_0).$$

Định lý 3. [9] *Lớp các ngôn ngữ được sinh bởi các văn phạm loại i trùng với lớp các ngôn ngữ được đoán nhận bởi các otomat loại i với* $i = 0, 1, 2, 3$.

3. HỌ TRỪU TƯỢNG CÁC NGUỒN

Định nghĩa 8. Họ trừu tượng các nguồn là một bộ hai $(\Omega, \mathcal{S})$ hay đơn giản là $\mathcal{S}$ khi Ω đã xác định và được xác định như sau:

1. Ω là một bộ sáu $(K, \Sigma, \Gamma, I, f, g)$ (đã được xác định ở Mục 2.4).

2. $\mathcal{S}$ là tập các nguồn $S = (T_S, V, E, H, \psi, v_0, F_S)$, trong đó:

2.1. T_S là tập con hữu hạn của Σ và thỏa mãn điều kiện sau:

$$T_S \cap (\Sigma_{1S} \cup \Sigma_{2S}) = \Phi \text{ với } \Sigma_{1S}, \Sigma_{2S} \text{ là các tập con của } \Sigma, \Sigma_{1S} \cap \Sigma_{2S} = \Phi,$$

2.2. V là tập các đỉnh,

2.3. E là tập các cung,
2.4. $H = \{h_1, h_2, h_T, h'_T, h_V, h_E\}$ là tập các phép ánh xạ và đồng cấu được xác định như sau:
2.4.1. h_1 là song ánh từ tập I_{1S} vào tập Σ_{1S} với I_{1S} là tập con hữu hạn của I,
2.4.2. h_2 là song ánh từ tập G_S vào tập Σ_{2S} với G_S là tập con hữu hạn của Γ^* và $G_S \subseteq g(\Gamma^*)$,
2.4.3. h_T là phép đồng cấu trên tập $(T_S \cup \Sigma_{1S} \cup \Sigma_{2S})^*$, sao cho

$$h_T(a) = \begin{cases} \varepsilon \ \text{ với mọi } a \in T_S, \\ a \text{ với mọi } a \in \Sigma_{1S} \cup \Sigma_{2S}, \end{cases}$$

2.4.4. h'_T là phép đồng cấu trên tập $(T_S \cup \Sigma_{1S} \cup \Sigma_{2S})^*$, sao cho

$$h_T(a)' = \begin{cases} a \ \text{ với mọi } a \in T_S, \\ \varepsilon \text{ với mọi } \ a \in \Sigma_{1S} \cup \Sigma_{2S}, \end{cases}$$

2.4.5. h_V là song ánh từ tập V vào tập $K_{1S} \times G_S$ và được gọi là hàm gán nhãn đỉnh với K_{1S} là tập con hữu hạn của K, h_V^{-1} là ánh xạ ngược của h_V,
2.4.6. h_E là song ánh từ tập E vào tập $(T_S \cup \{\varepsilon\})\Sigma_{1S} \times K_{1S}$ và được gọi là hàm gán nhãn cung, h_E^{-1} là ánh xạ ngược của h_E,
2.5. ψ là ánh xạ từ tập K_{1S} vào tập các tập con hữu hạn của $(T_S \cup \{\varepsilon\})\Sigma_{2S}\Sigma_{1S} \times K_{1S}$,
2.6. F_S là tập con của K_{1S},
2.7. $v_0 \in V$ là đỉnh ban đầu với nhãn $h_V(v_0) = (q_0, \varepsilon)$ và $V_F \subseteq V$ là tập các đỉnh kết thúc với các nhãn $h_V(V_F) = (F_S, \{\varepsilon\})$.

Để hình thức hoá việc hoạt động của nguồn S ta có khái niệm phần phải nhớ.

Phần phải nhớ được ghi tại các đỉnh của nguồn S là một bộ hai:

$$m = (q, \beta) \in K_{1S} \times \Gamma^*,$$

trong đó:

$q \in K_{1S}$ là trạng thái tức thời của nguồn S,

$\beta \in \Gamma^*$ là nội dung của băng làm việc trong nguồn S.

Giả sử tại đỉnh $v \in V$ của nguồn S ta có:

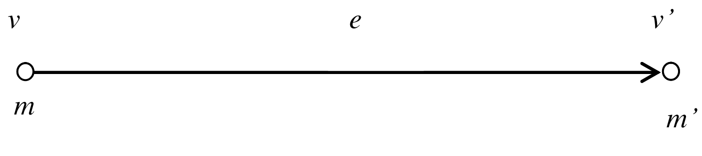
$$m = (q, \beta),\ \psi = q \to ah_2(\gamma)h_1(u)q' \ \text{ với điều kiện } \gamma \in g(\beta),\ h_V(v) = (q, \gamma).$$

Khi đó tại đỉnh $v' \in V$ của nguồn S, phần phải nhớ m' và nhãn của đỉnh v' được xác định như sau:

$m' = (q', f(\gamma, u))$,

$h_V(v') = (q', \gamma')$ với điều kiện $\gamma' \in g(f(\gamma, u))$.

Cung π với nhãn $e = ah_2(\gamma)h_1(u)$ nối từ đỉnh v với phần phải nhớ m tới đỉnh v' với phần phải nhớ m' được biểu diễn như sau:



và được viết bởi $(v, m)e(v', m')$.

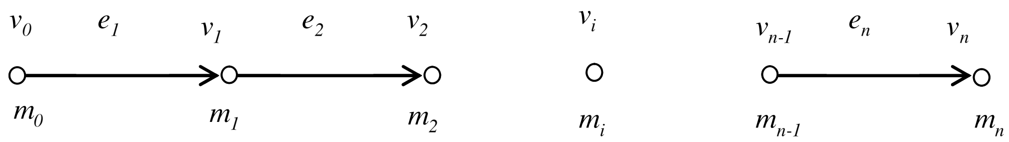
Đường Π độ dài n trong nguồn S đi từ đỉnh v_0 với phần phải nhớ m_0 tới đỉnh v_n với phần phải nhớ m_n là một dãy liên tiếp các cung $\pi_1, \pi_2, ..., \pi_n$, sao cho đỉnh cuối của cung π_i là đỉnh đầu của π_{i+1} với $1 \leqslant i \leqslant n-1$.

Nhãn λ của đường Π trong nguồn S được xác định bởi

$$h_E(\pi_1\pi_2...\pi_n) = e_1e_2...e_n.$$

Độ dài của đường Π được ký hiệu là $|\Pi|$.

Đường Π trong nguồn S được biểu diễn như sau:



trong đó:

$a_1, a_2, ..., a_n \in T_S \cup \{\varepsilon\}$,

$m_0 = (q_0, \beta_0)$,

$m_i = (q_i, \beta_i)$ với $\beta_i = f(\gamma_{i-1}, u_i),\ 1 \leqslant i \leqslant n-1$,

$\psi_1 = q_0 \to a_1h_2(\gamma_0)h_1(u_1)q_1$,

$\psi_{i+1} = q_i \to a_{i+1}h_2(\gamma_i)h_1(u_{i+1})q_{i+1}$ với $\gamma_i \in g(\beta_i), 1 \leqslant i \leqslant n-1$,

$h_V(v_0) = (q_0, \gamma_0)$,

$h_V(v_i) = (q_i, \gamma_i)$ với $1 \leqslant i \leqslant n-1$,

$h_V(v_n) = (q_n, \gamma_n)$ với $\gamma_n \in g(\beta_n)$,

$e_1 = a_1h_2(\gamma_0)h_1(u_1)$,

$e_{i+1} = a_{i+1}h_2(\gamma_i)h_1(u_{i+1})$ với $1 \leqslant i \leqslant n-1$,

và được viết bởi $(v_0, m_0)\lambda(v_n, m_n)$ với nhãn $\lambda = e_1e_2...e_n$.

Khi đó $\beta_i = f(\gamma_{i-1}, u_i)$ là phần tử thứ hai trong phần phải nhớ m_i với $1 \leqslant i \leqslant n-1$ có thể được xác định như sau:

$\beta_i = \Phi^i(\gamma_0, u_1, u_2, ..., u_i)$ với $1 \leqslant i \leqslant n-1, n$ là độ dài của đường trong nguồn S, Φ^n là hàm được xác định trên tích Đềcác $\Gamma^* \times I_{1S} \times ... \times I_{1S}(n$ lần) và thỏa mãn các điều kiện sau:

$\Phi^0(\gamma) = \gamma$ với $\gamma \in \Gamma^*$,

$\Phi^i(\gamma, u_1, u_2, ..., u_i) = f(\Phi^{i-1}(\gamma, u_1, u_2, ..., u_{i-1}), u_i)$ với $u_n \in I_{1S}, \gamma \in \Gamma^*, 1 \leqslant i \leqslant n$.

Đường trong nguồn S được gọi là đường đơn nếu đường đó đi qua mỗi cung không quá một lần.

Đường trong nguồn S được gọi là chu trình nếu đường đó bắt đầu và kết thúc tại cùng một đỉnh.

Chu trình trong nguồn S được là chu trình đơn nếu chu trình đó đi qua mỗi cung không quá một lần.

Đường Π trong nguồn S được gọi là đường đầy đủ với nhãn λ nếu đường đó đi từ đỉnh ban đầu v_0 với phần phải nhớ ban đầu m_0 tới đỉnh kết thúc v_f với phần phải nhớ kết thúc m_f và được viết bởi $(v_0, m_0)\lambda(v_f, m_f)$, trong đó:

$m_0 = (q_0, \varepsilon)$,

$m_i = (q_i, \beta_i)$ với $\beta_i = \Phi^i(\varepsilon, u_1, u_2, ..., u_i), 1 \leqslant i \leqslant n-1$,

$m_f = (q_f, \beta_n) = (q_f, \varepsilon)$ với $\beta_n = \Phi^n(\varepsilon, u_1, u_2, ..., u_n) = \varepsilon$.

Khi đó:

$M_S = (m_0, m_1, ..., m_f)$ là dãy phải nhớ đầy đủ của đường đầy đủ Π trong nguồn S,

$t = h_T(\lambda) = h_2(\varepsilon)h_1(u_1)...h_2(\gamma_{n-1})h_1(u_n) \in \Sigma_{1S} \cup (\Sigma_{2S})^*$ là vết của đường đầy đủ Π trong nguồn S,

$w = h'_T(\lambda) = a_1a_2...a_n \in T_S^*$ là xâu (hay từ) được xác định bởi đường đầy đủ Π trong nguồn S.

Định nghĩa 9. Vết của nguồn S là tập:

$$L_t = \{t \in (\Sigma_{1S} \cup \Sigma_{2S})^* \mid t = h_T(\lambda),\ (v_0, m_0)\lambda(v_f, m_f)\}.$$

Ngôn ngữ được xác định bởi nguồn S là tập:

$$L(S) = \{w \in T_S^* \mid h_T(\lambda) = t \in L_t, h'_T(\lambda) = w, (v_0, m_0)\lambda(v_f, m_f)\}.$$

Họ các ngôn ngữ được xác định bởi họ trừu tượng các nguồn là tập:

$$\mathcal{L}(\mathcal{S}) = \{L(S) \mid S \in \mathcal{S}\}.$$

Định lý 4. *Nếu $L_t \subseteq (\Sigma_{1S} \cup \Sigma_{2S})^*$ là vết của nguồn S và $h_2(\varepsilon) \in \Sigma_{2S}$ thì $L_t h_2(\varepsilon)$ được đoán nhận bởi otomat $A = (K_A, \Sigma_{1S} \cup \Sigma_{2S}, \delta, q_0, F_A) \in \mathcal{A}$.*

Chứng minh:

Otomat $A = (K_A, \Sigma_{1S} \cup \Sigma_{2S}, \delta, q_0, F_A) \in \mathcal{A}$ được xác định như sau:

$K_A = \{q_0, q_1, q_2\}$, $F_A = \{q_1\}$, $\delta(q_0, h_2(\varepsilon), \varepsilon) = \{(q_1, 1_\varepsilon)\}$, $\delta(q_1, h_1(u), \gamma) = \{(q_2, u)\}$ với $\gamma \in g(\Gamma^*)$ và $u \in I_{1S}$, $\delta(q_2, h_2(\gamma), \gamma) = \{(q_1, 1_\gamma)\}$ với $\gamma \in g(\Gamma^*)$.

Ta có quan hệ sau: $(q_0, th_2(\varepsilon), \varepsilon) \vdash^* (q_1, \varepsilon, \varepsilon)$ với $th_2(\varepsilon) = h_2(\varepsilon)h_1(u)...h_2(\gamma_{n-1})h_1(u_n)h_2(\varepsilon) \in L_t h_2(\varepsilon)$.

Vậy $L_t h_2(\varepsilon)$ được đoán nhận bởi otomat $A \in \mathcal{A}$.

Định lý đã được chứng minh. ■

Nhận xét 2

- Do K_{1S}, T_S, I_{1S}, G_S là các tập hữu hạn nên tập Σ_{1S}, tập Σ_{2S}, tập các đỉnh V và tập các cung E cũng là hữu hạn. Vì vậy nguồn S được xem như là đồ thị hữu hạn có hướng được gán nhãn và tại các đỉnh có ghi các phần phải nhớ.

- Đường đầy đủ Π trong nguồn S với nhãn λ được cấu thành từ đường đầy đủ đơn và các chu trình đơn (nếu có).

- Đường đầy đủ Π trong nguồn S với nhãn λ được viết bởi:

$$(v_0, m_0)w/t(v_f, m_f) \text{ với } t = h_T(\lambda) \text{ và } w = h'_T(\lambda).$$

- Ngôn ngữ được xác định bởi nguồn S là tập:

$$L(S) = \{w \in T_S^* \mid w = h'_T(\lambda), (v_0, m_0)w/t(v_f, m_f) \text{ với } t \in L_t\}.$$

Bổ đề 1. *Đối với nguồn $S \in \mathcal{S}$, luôn tồn tại otomat $A \in \mathcal{A}$, sao cho $L(A) = L(S)$.*

Chứng minh:

Giả sử $S = (T_S, V, E, H, \psi, v_0, F_S) \in \mathcal{S}$ là nguồn.

Otomat $A = (K_1, T, \delta, q_0, F) \in \mathcal{A}$ được xây dựng như sau:

$K_1 = K_{1S}, T = T_S, I_1 = I_{1S}, G_A = G_S$,

$(q', u) \in \delta(q, a, \gamma)$ với $\psi = q \to ah_2(\gamma)h_1(u)q'$,

$q_0 \in K_1$ là trạng thái ban đầu,

$F = F_S$ là tập các trạng thái kết thúc.

Ta cần chứng minh $L(A) = L(S)$.

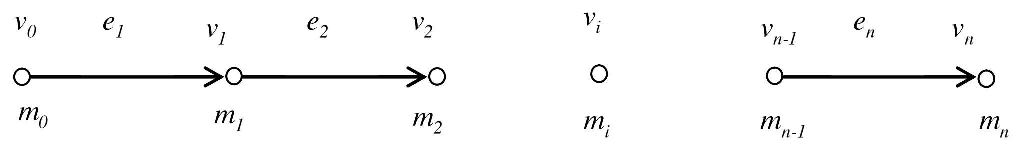
Trước tiên ta cần chứng minh bao hàm thức $L(S) \subseteq L(A)$.

Xét xâu bất kỳ khác rỗng $w \in L(S)$.

Do $w \in L(S)$ nên tồn tại đường đầy đủ Π trong nguồn S với nhãn $\lambda = e_1 e_2 ... e_n$ và dãy phải nhớ đầy đủ $M_S = (m_0, m_1, ..., m_f)$, sao cho

$$h_T(\lambda) = t \in L_t, h'_T(\lambda) = a_1 a_2 ... a_n = w.$$

Đường đầy đủ Π trong nguồn S được biểu diễn như sau:



trong đó:

$a_1, a_2, ..., a_n \in T_S \cup \{\varepsilon\}$,
$m_0 = (q_0, \beta_0)$,
$m_i = (q_i, \beta_i)$ với $\beta_i = f(\gamma_{i-1}, u_i)$, $1 \leqslant i \leqslant n-1$,
$\psi_1 = q_0 \to a_1 h_2(\gamma_0) h_1(u_1) q_1$,
$\psi_{i+1} = q_i \to a_{i+1} h_2(\gamma_i) h_1(u_{i+1}) q_{i+1}$ với $\gamma_i \in g(\beta_i), 1 \leqslant i \leqslant n-1$,
$h_V(v_0) = (q_0, \gamma_0)$,
$h_V(v_i) = (q_i, \gamma_i)$ với $1 \leqslant i \leqslant n-1$,
$h_V(v_f) = (q_n, \gamma_i)$,
$e_1 = a_1 h_2(\gamma_0) h_1(u_1)$,
$e_{i+1} = a_{i+1} h_2(\gamma_i) h_1(u_{i+1})$ với $1 \leqslant i \leqslant n-1$.

Khi đó ta có:

$a_1, a_2, ..., a_n \in T \cup \{\varepsilon\}$,
q_0 là trạng thái ban đầu,
$q_1, q_2, ..., q_{n-1} \in K_1$,
$q_n \in F$ là trạng thái kết thúc,
$\beta_i = f(\Phi^{i-1}, u_i)$ với $1 \leqslant i \leqslant n-1$,
$(q_1, u_1) \in \delta(q_0, a_1, \varepsilon)$ được xác định từ ψ_1,
$(q_{i+1}, u_{i+1}) \in \delta(q_i, a_{i+1}, \gamma_i)$ được xác định từ ψ_{i+1} với $\gamma_i \in g(\beta_i)$, $1 \leqslant i \leqslant n-1$, và dãy đầy đủ các bước chuyển C của otomat A được biểu diễn như sau:

$$\begin{aligned}(q_0, a_1 a_2 ... a_n, \varepsilon) &\vdash (q_1, a_2 ... a_n, \beta_1)\\ &\dots\\ &\vdash (q_{n-1}, a_n, \beta_{n-1})\\ &\vdash (q_n, \varepsilon, \varepsilon). \end{aligned} \tag{1}$$

Suy ra $w = a_1 a_2 ... a_n \in L(A)$. Vậy $L(S) \subseteq L(A)$.

Ngược lại, ta cần chứng minh bao hàm thức $L(A) \subseteq L(S)$.

Xét xâu bất kỳ khác rỗng $w \in L(A)$.

Do $w \in L(A)$ nên tồn tại dãy các hàm chuyển và dãy đầy đủ các bước chuyển C của otomat A, sao cho $w = a_1 a_2 ... a_n$.

Dãy đầy đủ của các bước chuyển C của otomat A được biểu diễn như sau:

$$
\begin{aligned}
(q_0, a_1a_2...a_n, \varepsilon) &\vdash (q_1, a_1...a_n, \beta_1) \\
&\dots \\
&\vdash (q_{n-1}, a_n, \beta_{n-1}) \\
&\vdash (q_n, \varepsilon, \varepsilon). \qquad (2)
\end{aligned}
$$

trong đó: $a_1, a_2, ..., a_n \in T \cup \{\varepsilon\}$,
q_0 là trạng thái ban đầu,
$q_1, q_2, ..., q_{n-1} \in K_1$,
$q_n \in F$ là trạng thái kết thúc,
$\beta_i = f(\Phi^{i-1}, u_i)$ với $1 \leqslant i \leqslant n-1$,
$(q_1, u_1) \in \delta(q_0, a_1, \varepsilon)$,
$(q_{i+1}, u_{i+1}) \in \delta(q_i, a_{i+1}, \gamma_i)$ với $\gamma_i \in g(\beta_i), 1 \leqslant i \leqslant n-1$.
Khi đó ta có:
$a_1, a_2, ..., a_n \in T_S \cup \{\varepsilon\}$,
$m_0 = (q_0, \varepsilon)$,
$m_i = (q_i, \Phi^i)$ với $1 \leqslant i \leqslant n-1$,
$m_f = (q_n, \varepsilon)$,
$\psi_1 = q_0 \to a_1h_2(\varepsilon)h_1(u_1)q_1$ được xác định từ $(q_1, u_1) \in \delta(q_0, a_1, \varepsilon)$,
$\psi_{i+1} = q_i \to a_{i+1}h_2(\gamma_i)h_1(u_{i+1})q_{i+1}$ được xác định từ $(q_{i+1}, u_{i+1}) \in \delta(q_i, a_{i+1}, \gamma_i)$ với $\gamma_i \in g(\Phi^i)$, $1 \leqslant i \leqslant n-1$,
$h_V^{-1}(q_0, \varepsilon) = v_0$,
$h_V^{-1}(q_i, \gamma_i) = v_i$ với $1 \leqslant i \leqslant n-1$,
$h_V^{-1}(q_n, \varepsilon) = v_f$,
$e_1 = a_1h_2(\varepsilon)h_1(u_1)$,
$e_{i+1} = a_{i+1}h_2(\gamma_i)h_1(u_{i+1})$ với $1 \leqslant i \leqslant n-1$,
và đường đầy đủ Π trong nguồn S với nhãn $\lambda = e_1e_2...e_n$ được biểu diễn như sau:

Suy ra
$M_S = (m_0, m_1, ..., m_f)$,
$h_T(\lambda) = t \in L_t$,
$h'_T(\lambda) = a_1a_2...a_n = w \in L(S)$.
Vậy $L(A) \subseteq L(S)$.
Kết luận $L(A) = L(S)$. ■

Bổ đề 2. *Đối với otomat $A \in \mathcal{A}$ luôn tồn tại nguồn $S \in \mathcal{S}$, sao cho $L(S) = L(A)$.*

Chứng minh:
Chứng minh $L(S) = L(A)$ tương tự như ở Bổ đề 1.

Từ Bổ đề 1 và Bổ đề 2 ta có định lý sau:

Định lý 5. *Họ các ngôn ngữ được xác định bởi họ trừu tượng các nguồn trùng với họ các ngôn ngữ được đoán nhận bởi họ trừu tượng các otomat.*

Nhận xét 3

Tập các dãy đầy đủ bước chuyển trong otomat và tập các đường đầy đủ trong nguồn tương ứng có sự tương ứng 1–1.

4. MỘT SỐ LỚP NGUỒN

Trên cơ sở định nghĩa họ trừu tượng các nguồn và dựa vào đặc điểm của xâu trong phần phải nhớ, ta tách từ họ trừu tượng các nguồn ra một số lớp:

Lớp các nguồn loại 0

Nguồn được gọi là nguồn loại 0, nếu không có sự hạn chế và ràng buộc nào đối với xâu trong phần phải nhớ.

Lớp các nguồn loại 0, ký hiệu bởi $\mathcal{S}_0$, là tập các nguồn loại 0.

Lớp các ngôn ngữ được xác định bởi các nguồn loại 0, ký hiệu bởi $\mathcal{L}(\mathcal{S}_0)$, là tập các ngôn ngữ được xác định bởi các nguồn loại 0.

Lớp các nguồn loại 1

Nguồn được gọi là nguồn loại 1, nếu độ dài của xâu trong phần phải nhớ luôn bị giới hạn bởi hàm tuyến tính với biến là độ dài của xâu vào.

Lớp các nguồn loại 1, ký hiệu bởi $\mathcal{S}_1$, là tập các nguồn loại 1.

Lớp các ngôn ngữ được xác định bởi các nguồn loại 1, ký hiệu bởi $\mathcal{L}(\mathcal{S}_1)$, là tập các ngôn ngữ được xác định bởi các nguồn loại 1.

Lớp các nguồn loại 2

Nguồn được gọi là nguồn loại 2, nếu quá trình biến đổi nội dung của xâu trong phần phải nhớ luôn được thực hiện ở phần bên phải nhất (hoặc bên trái nhất).

Lớp các nguồn loại 2, ký hiệu bởi $\mathcal{S}_2$, là tập các nguồn loại 2.

Lớp các ngôn ngữ được xác định bởi các nguồn loại 2, ký hiệu bởi $\mathcal{L}(\mathcal{S}_2)$, là tập các ngôn ngữ được xác định bởi các nguồn loại 2.

Lớp các nguồn loại 3

Nguồn được gọi là nguồn loại 3, nếu xâu trong phần phải nhớ luôn là xâu rỗng.

Lớp các nguồn loại 3, ký hiệu bởi $\mathcal{S}_3$, là tập các nguồn loại 3.

Lớp các ngôn ngữ được xác định bởi các nguồn loại 3, ký hiệu bởi $\mathcal{L}(\mathcal{S}_3)$, là tập các ngôn ngữ được xác định bởi các nguồn loại 3.

Lớp các ngôn ngữ được xác định bởi các nguồn tương ứng với sự phân lớp trên thỏa mãn bao hàm thức sau:

$$\mathcal{L}(\mathcal{S}_3) \subset \mathcal{L}(\mathcal{S}_2) \subset \mathcal{L}(\mathcal{S}_1) \subset \mathcal{L}(\mathcal{S}_0).$$

Định lý 6. *Lớp các ngôn ngữ được xác định bởi các nguồn loại i trùng với lớp các ngôn ngữ được đoán nhận bởi các otomat loại i với* $i = 0, 1, 2, 3$.

Chứng minh:

Dựa trên định nghĩa họ trừu tượng các otomat, họ trừu tượng các nguồn và từ sự phân lớp otomat, sự phân lớp nguồn ta nhận thấy:

Quá trình biến đổi nội dung xâu trong phần phải nhớ của các nguồn loại 0, nguồn loại 1, nguồn loại 2, nguồn loại 3 được thực hiện tương tự như trên băng làm việc của các otomat loại 0, otomat loại 1, otomat loại 2, otomat loại 3 và sự phân lớp nguồn hoàn toàn giống với sự phân lớp otomat.

Kết hợp nhận xét trên với Định lý 5 ta có:

Lớp các ngôn ngữ được xác định bởi các nguồn loại i trùng với lớp các ngôn ngữ được đoán nhận bởi các otomat loại i với $i = 0, 1, 2, 3$.

Định lý đã được chứng minh. ■

Định lý 7. *Nếu $\mathcal{S}$ là họ trừu tượng các nguồn thì $\mathcal{L}(\mathcal{S})$ là họ trừu tượng đầy đủ các ngôn ngữ.*

Định lý 8. *Nếu T là hàm tăng thì $\mathcal{L}_{T(n)}(\mathcal{S})$ là họ trừu tượng các ngôn ngữ.*

Chứng minh Định lý 7 và Định lý 8:

Từ Định lý 1 ta có:

Nếu $\mathcal{A}$ là họ trừu tượng các otomat thì $\mathcal{L}(\mathcal{A})$ là họ trừu tượng đầy đủ các ngôn ngữ.

Từ Định lý 2 ta có:

Nếu T là hàm tăng thì $\mathcal{L}_{T(n)}(\mathcal{A})$ là họ trừu tượng các ngôn ngữ.

Do họ trừu tượng các nguồn $\mathcal{S}$ tương đương với họ trừu tượng các otomat $\mathcal{A}$ trong việc xác định các ngôn ngữ và sự phân lớp nguồn hoàn toàn giống với sự phân lớp otomat nên ta có:

$\mathcal{L}(\mathcal{S})$ là họ trừu tượng đầy đủ các ngôn ngữ và $\mathcal{L}_{T(n)}(\mathcal{S})$ là họ trừu tượng các ngôn ngữ.

Định lý 7 và Định lý 8 đã được chứng minh. ■

Định lý 9. *Các khẳng định sau đây là tương đương:*

1) *$\mathcal{L}_i = \mathcal{L}(\mathcal{G}_i)$ là lớp các ngôn ngữ được sinh bởi các văn phạm loại i.*

2) *$\mathcal{L}_i = \mathcal{L}(\mathcal{A}_i)$ là lớp các ngôn ngữ được đoán nhận bởi các otomat loại i.*

3) *$\mathcal{L}_i = \mathcal{L}(\mathcal{S}_i)$ là lớp các ngôn ngữ được xác định bởi các nguồn mở rộng loại i với $i = 0, 1, 2, 3$.*

Chứng minh:

i) Từ Định lý 3 ta có: 2) được suy ra từ 1) và ngược lại 1) được suy ra từ 2),

ii) Từ Định lý 6 ta có: 3) được suy ra từ 2) và ngược lại 2) được suy ra từ 3),

iii) Kết hợp ii) với i) ta có: 3) được suy ra từ 1) và ngược lại 1) được suy ra từ 3).

Định lý đã được chứng minh. ■

5. KẾT LUẬN

Họ các ngôn ngữ được xác định bởi họ trừu tượng các nguồn trùng với họ các ngôn ngữ được đoán nhận bởi họ trừu tượng các otomat.

Đặc biệt lớp các ngôn ngữ được sinh bởi các văn phạm loại i, lớp các ngôn ngữ được đoán nhận bởi các otomat loại i và lớp các ngôn ngữ được xác định bởi các nguồn loại i là trùng nhau với $i = 0, 1, 2, 3$.

Với các công cụ đại số và hình học (cụ thể là đồ thị) họ trừu tượng các nguồn làm phong phú thêm lý thuyết ngôn ngữ hình thức, otomat và ứng dụng rộng rãi trong những lĩnh vực khác nhau của tin học: chương trình dịch, độ phức tạp tính toán của các máy, tính toán song song...

TÀI LIỆU THAM KHẢO

[1] Phùng Văn Ổn, "Nghiên cứu một số lớp siêu ngôn ngữ", Luận án tiến sĩ toán học, Hà Nội, 2001.

[2] Đặng Huy Ruận, *Lý thuyết ngôn ngữ hình thức và otomat,* NXB Đại học Quốc gia Hà Nội, 2002.

[3] A. V. Aho, J. D. Ullman, The theory of languages, *Math. Syst. Theory* **2** (1968) 97–125.

[4] N. Chomsky, On certain formal properties of grammars, *Information and Control* **2** (1959) 137–167.

[5] S. Ginsburg, S.Greibach, Abstract familles of languages, *Mem. Amer. Math. Soc.* **87** (1969) 1–32.

[6] T. Nipkow (Ed.) *Rewriting Techniques and Applications, Lecture Notes in Computer Science* Springer Verlag, Berlin, 1998.

[7] T. A. Sudkamd, *Languages and Machines,* Addison-Wesley, Reading Massachusetts, 1997.

[8] R. G. Taylor, *Models of Computation and Formal Languages,* Oxford University Press, New York, 1998.

[9] A. B. Geadkii, *Văn phạm hình thức và ngôn ngữ*, Nauka, Mockva, 1973 (tiếng Nga).

Nhận bài ngày 16 - 3 - 2004

Nhận lại sau sửa ngày 5 - 5 - 2006