Tạp chí Tin học và Điều khiển học, T.21, S.4 (2005), 301–311

ĐỒ HÌNH TỔNG QUÁT VÀ MỐI QUAN HỆ VỚI VĂN PHẠM CẤU TRÚC - CÂU

NGUYỄN THẾ BÌNH

Đại học Dân lập Hải Phòng

Abstract. This article gives the basic concepts of general diagram and proves that the family of languages defined by general diagram coincides with the family of languages generated by phrase-structure grammar.

Tóm tắt. Bài báo đưa ra những khái niệm cơ bản về đồ hình tổng quát và chứng minh rằng lớp các ngôn ngữ được xác định bởi đồ hình tổng quát trùng với lớp các ngôn ngữ được sinh bởi văn phạm cấu trúc - câu.

1. ĐẶT VẤN ĐỀ

Ngôn ngữ hình thức và ôtômat là cơ sở quan trọng của tin học, một ngành khoa học và công nghệ đang phát triển mạnh mẽ theo nhiều hướng khác nhau [1–11]. Một trong những hướng phát triển đó là xây dựng mô hình đồ hình, thiết lập mối quan hệ đồ hình với văn phạm chính quy được đưa ra trong [11], tiếp tục được nghiên cứu trong [5, 6] và mở rộng khái niệm đồ hình thành ω - sơ đồ sinh, làm công cụ sinh ra lớp ω - ngôn ngữ chính quy, xác định độ phức tạp đoán nhận của ôtômat hữu hạn đối với lớp ω - ngôn ngữ chính quy đó được đưa ra trong [7, 8].

Bài báo này phát triển khái niệm đồ hình theo hướng tổng quát hóa như mở rộng khái niệm đồ hình thành đồ hình tổng quát, phân loại đồ hình tổng quát, thiết lập mối quan hệ đồ hình tổng quát với văn phạm cấu trúc - câu và chứng minh rằng lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát trùng với lớp các ngôn ngữ được sinh bởi các văn phạm cấu trúc - câu.

2. CÁC KHÁI NIỆM CƠ BẢN

2.1. Văn phạm và ngôn ngữ được sinh bởi văn phạm cấu trúc - câu ([1, 6, 8, 11])

Định nghĩa 1. Văn phạm cấu trúc - câu, hay gọn hơn văn phạm, là một bộ bốn: $G = (N, \Sigma, P, \sigma)$, trong đó:
N là tập hữu hạn các ký hiệu không kết thúc và được gọi là bảng chữ cái phụ,
Σ là tập hữu hạn các ký hiệu kết thúc và được gọi là bảng chữ cái chính,
$N \cap \Sigma = \emptyset$,
$N \cup \Sigma = V$ là tập hỗn hợp hay là bảng chữ cái đầy đủ,
$\sigma \in N$ là ký hiệu ban đầu,

P là tập hữu hạn các quy tắc có dạng sau:

$$\varphi \to \psi \text{ với } \varphi, \psi \in V^* \quad \text{và} \quad \to \notin V.$$

Bước dẫn trong văn phạm G trên V^* được xác định bởi quan hệ hai ngôi $\Rightarrow$:

$$\alpha \Rightarrow \beta \text{ với } \alpha = \xi\varphi\eta;\ \beta = \xi\psi\eta;\ \varphi \to \psi \in P;\ \alpha, \beta, \xi, \eta, \psi \in V^*.$$

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

Định nghĩa 2. Ngôn ngữ được sinh bởi văn phạm cấu trúc - câu G là tập:

$$L(G) = \{\omega \in \Sigma^* | \sigma \Rightarrow^* \omega\}.$$

Định nghĩa 3. Giả sử $\mathcal{G}$ là lớp các văn phạm.

Lớp các ngôn ngữ được sinh bởi các văn phạm thuộc $\mathcal{G}$ là tập:

$$\mathcal{L}(\mathcal{G}) = \{L(G) | G \in \mathcal{G}\}.$$

Trên cơ sở định nghĩa văn phạm cấu trúc - câu, N. Chomsky đã đưa ra sự phân loại văn phạm như sau:

Văn phạm loại 0

Văn phạm cấu trúc - câu được gọi là văn phạm loại 0, nếu mọi quy tắc trong P đều có đạng $\varphi \to \psi$ với $\varphi, \psi \in V^*$ và không có sự hạn chế và ràng buộc nào.

Văn phạm loại 0 còn được gọi là văn phạm ngữ cấu.

Ngôn ngữ được sinh bởi văn phạm loại 0 được gọi là ngôn ngữ ngữ cấu.

Lớp các văn phạm loại o, ký hiệu bởi $\mathcal{G}_o$, được là tập các văn phạm loại 0.

Lớp các ngôn ngữ được sinh bởi các văn phạm loại 0, được ký hiệu bởi $\mathcal{L}(\mathcal{G}_o)$, là tập các ngôn ngữ ngữ cấu.

Văn phạm loại 1

Văn phạm cấu trúc - câu được gọi là văn phạm loại 1, nếu mọi quy tắc trong P đều có đạng $\varphi \to \psi$ với $|\varphi| \leqslant |\psi|;\ \varphi, \psi \in V^*$.

Văn phạm loại 1 còn được gọi là văn phạm cảm ngữ cảnh.

Ngôn ngữ được sinh bởi văn phạm loại 1 được gọi là ngôn ngữ cảm ngữ cảnh.

Lớp các văn phạm loại 1, ký hiệu bởi $\mathcal{G}_1$, là tập các văn phạm loại 1.

Lớp các ngôn ngữ được sinh bởi các văn phạm loại 1, được ký hiệu bởi $\mathcal{L}(\mathcal{G}_1)$, là tập các ngôn ngữ cảm ngữ cảnh.

Văn phạm loại 2

Văn phạm cấu trúc - câu được gọi là văn phạm loại 2, nếu mọi quy tắc trong P đều có đạng $A \to \alpha$ với $A \in N, \alpha \in V^*$.

Văn phạm loại 2 còn được gọi là văn phạm phi ngữ cảnh.

Ngôn ngữ được sinh bởi văn phạm loại 2 được gọi là ngôn ngữ phi ngữ cảnh.

Lớp các văn phạm loại 2, được ký hiệu bởi $\mathcal{G}_2$, là tập các văn phạm loại 2.

Lớp các ngôn ngữ được sinh bởi các văn phạm loại 2, được ký hiệu bởi $\mathcal{L}(\mathcal{G}_2)$, là tập các ngôn ngữ phi ngữ cảnh.

Văn phạm loại 3

Văn phạm cấu trúc - câu được gọi là văn phạm loại 3, nếu mọi quy tắc trong P thuộc một trong hai dạng sau:

i) $A \to aB$, ii) $A \to a$, với $A, B \in N, a \in \Sigma$.

Văn phạm loại 3 còn được gọi là văn phạm chính quy.

Ngôn ngữ được sinh bởi văn phạm loại 3 được gọi là ngôn ngữ chính quy.

Lớp các văn phạm loại 3, được ký hiệu bởi $\mathcal{G}_3$, là tập các văn phạm loại 3.

Lớp các ngôn ngữ được sinh bởi các văn phạm loại 3, được ký hiệu bởi $\mathcal{L}(\mathcal{G}_3)$, là tập các ngôn ngữ chính quy.

Lớp các ngôn ngữ được sinh bởi các văn phạm tương ứng với sự phân loại trên thỏa mãn bao hàm thức sau:

$$\mathcal{L}(\mathcal{G}_3) \subset \mathcal{L}(\mathcal{G}_2) \subset \mathcal{L}(\mathcal{G}_1) \subset \mathcal{L}(\mathcal{G}_o).$$

2.2. Sự tương đương giữa các ôtômat và các văn phạm cấu trúc - câu ([4, 9, 10, 11])

Định nghĩa 4. Ta nói rằng ôtômat A và văn phạm G là tương đương nếu $L(A) = L(G)$.

Định lý 1.

i) *Đối với máy Turing bất kỳ luôn xây dựng được văn phạm ngữ cấu tương đương với nó.*

ii) *Đối với văn phạm ngữ cấu bất kỳ luôn xây dựng được máy Turing tương đương với nó.*

Định lý 2.

i) *Đối với ôtômat tuyến tính giới nội bất kỳ luôn xây dựng được văn phạm cảm ngữ cảnh tương đương với nó.*

ii) *Đối với văn phạm cảm ngữ cảnh bất kỳ luôn xây dựng được ôtômat tuyến tính giới nội tương đương với nó.*

Định lý 3.

i) *Đối với ôtômat đẩy xuống bất kỳ luôn xây dựng được văn phạm phi ngữ cảnh tương đương với nó.*

ii) *Đối với văn phạm phi ngữ cảnh bất kỳ luôn xây dựng được ôtômat đẩy xuống tương đương với nó.*

Định lý 4.

i) *Đối với ôtômat hữu hạn bất kỳ luôn xây dựng được văn phạm chính quy tương đương với nó.*

ii) *Đối với văn phạm chính quy bất kỳ luôn xây dựng được ôtômat hữu hạn tương đương với nó.*

Từ các định lý trên và từ sự phân loại văn phạm, sự phân lớp ôtômat [4] ta có định lý sau:

Định lý 5. *Lớp các ngôn ngữ được sinh bởi các văn phạm loại i trùng với lớp các ngôn ngữ được đoán nhận bởi các ôtômat loại i với $i = 0, 1, 2, 3$.*

3. ĐỒ HÌNH TỔNG QUÁT VÀ NGÔN NGỮ ĐƯỢC XÁC ĐỊNH BỞI ĐỒ HÌNH TỔNG QUÁT

Định nghĩa 5. Đồ hình tổng quát, hay gọn hơn, đồ hình là một bộ mười:

$$D = (\Sigma_N, \Sigma_N^-, \Sigma_T, \Sigma_T^-, W, V, E, H, v_0, v_f),$$

trong đó:

1) Σ_N, Σ_N^- là các tập hữu hạn các ký hiệu không kết thúc và có quan hệ như sau:

Đối với mỗi $A \in \Sigma_N$ luôn xác định được $A^- \in \Sigma_N^-$ sao cho $A^-A = \varepsilon$, ngược lại, đối với mỗi $A^- \in \Sigma_N^-$ luôn xác định được $A \in \Sigma_N$ sao cho $A^-A = \varepsilon$ và $\Sigma_N \cap \Sigma_N^- = \emptyset$.

2) Σ_T, Σ_T^- là các tập hữu hạn các ký hiệu kết thúc và có quan hệ như sau:

Đối với mỗi $a \in \Sigma_T$ luôn xác định được $a^- \in \Sigma_T^-$ sao cho $a^-a = \varepsilon$, ngược lại đối với mỗi $a^- \in \Sigma_T^-$ luôn xác định được $a \in \Sigma_T$ sao cho $a^-a = \varepsilon$ và

$$\Sigma_T \cap \Sigma_T^- = \emptyset,\ \Sigma_N \cap \Sigma_T = \emptyset, \Sigma_N^- \cap \Sigma_T^- = \emptyset,$$

$U = \Sigma_N \cup \Sigma_T$ là tập hỗn hợp các ký hiệu của Σ_N và Σ_T,
$U^- = \Sigma_N^- \cup \Sigma_T^-$ là tập hỗn hợp các ký hiệu của Σ_N^- và Σ_T^-,
U_V là tập con hữu hạn của U^*.

3) W là tập con hữu hạn của $(U)^*(U^-)^*$ và được gọi là tập các xâu ghép tiếp.

4) V là tập các đỉnh.

5) E là tập các cung.

6) $H = \{f, g, h, h_N, h_T, h_V, h_E\}$ là tập các phép ánh xạ và đồng cấu được xác định như sau:

6.1) f là ánh xạ từ tập $(U \cup U^-)^*$ vào tập $(U \cup U^-)^*$.

6.2) g là ánh xạ từ tập $(U \cup U^-)^*$ vào tập các tập con hữu hạn của $(U \cup U^-)^*$ và có tính chất sau: $g(\varepsilon) = \varepsilon$ và $\varepsilon \in g(\omega)$ khi và chỉ khi $\omega = \varepsilon$.

6.3) h là song ánh từ tập U^* vào tập $(U^-)^*$ và có tính chất sau: đối với mỗi $\omega \in U^*$ luôn xác định được $h(\omega) = \omega^- \in (U^-)^*$, sao cho $\omega^-\omega = \varepsilon$,

h^{-1} là ánh xạ ngược của h và có tính chất sau: đối với mỗi $\omega^- \in (U^-)^*$ luôn xác định được $h^{-1}(\omega^-) = \omega \in U^*$, sao cho $\omega^-\omega = \varepsilon$.

6.4) h_N là phép đồng cấu trên tập $(\Sigma_N \cup \Sigma_N^- \cup \Sigma_T \cup \Sigma_T^-)^*$, sao cho

$$h_N(a) = \begin{cases} \varepsilon & \text{với mọi } a \in (\Sigma_T \cup \Sigma_T^-), \\ a & \text{với mọi } a \in (\Sigma_N \cup \Sigma_N^-). \end{cases}$$

6.5) h_T là phép đồng cấu trên tập $(\Sigma_N \cup \Sigma_N^- \cup \Sigma_T \cup \Sigma_T^-)^*$, sao cho

$$h_T(a) = \begin{cases} a & \text{với mọi } a \in (\Sigma_T \cup \{\varepsilon\}), \\ \varepsilon & \text{với mọi } a \in (\Sigma_N \cup \Sigma_N^- \cup \Sigma_T^-)^*. \end{cases}$$

6.6) h_V là song ánh từ tập V vào tập U_V và được gọi là hàm gán nhãn đỉnh, $h^{-1}V$ là ánh xạ ngược của h_V.

6.7) h_E là song ánh từ tập E vào tập W và được gọi là hàm gán nhãn cung, $h^{-1}E$ là ánh xạ ngược của h_E.

7) $v_0 \in V$ là đỉnh ban đầu với nhãn $h_V(v_0) = \sigma_D$.

8) $v_f \in V$ là đỉnh kết thúc với nhãn $h_V(v_f) = K$ và thỏa mãn điều kiện $K \to \varepsilon$.

Để hình thức hóa việc hoạt động của đồ hình tổng quát D^- ta có khái niệm phần ghi nhớ.

Phần ghi nhớđược ký hiệu bởi μ, là một xâu thuộc $(U \cup U^-)^*$ và được ghi tại các đỉnh của đồ hình tổng quát D (ta quy ước phần đầu của μ được bố trí trên băng làm việc Ξ_D ở bên phải còn phần cuối của μ được bố trí trên băng làm việc Ξ_D ở bên trái).

Phần ghi nhớ được gọi là phần ghi nhớ rút gọn nếu sau khi biến đổi trên băng làm việc Ξ_D phần ghi nhớ đó chỉ gồm các ký hiệu thuộc tập U.

Giả sử tại đỉnh $v \in V$ của đồ hình tổng quát D ta có:

$$\mu = x\varphi y,$$

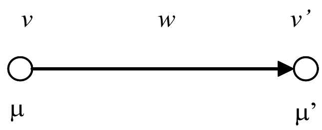
$$w = \psi\varphi^- = \psi h(\varphi),$$

$$h_V(v) = \varphi \in g(\mu).$$

Khi đó tại đỉnh $v' \in V$, phần ghi nhớ μ' và nhãn của đỉnh v' được xác định như sau:

$$\mu' = f(\mu, w) = f(x\varphi y, \psi\varphi^-) = x\psi\varphi^-\varphi y \approx x\psi y,\ h_V(v') = \varphi' \in g(\mu').$$

Cung e với nhãn w nối từ đỉnh v với phần ghi nhớ μ tới đỉnh v' với phần ghi nhớ μ' được biểu diễn như sau:



và được viết bởi $(v, \mu)w(v', \mu')$.

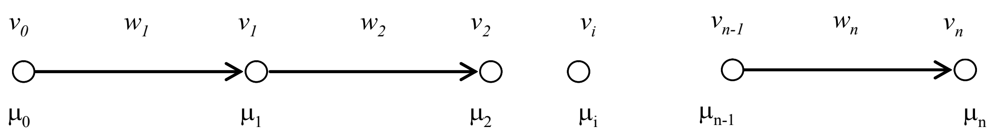
Đường Π độ dài n trong đồ hình tổng quát D đi từ đỉnh v_0 với phần ghi nhớ μ_0 tới đỉnh v_n với phần ghi nhớ μ_n là dãy liên tiếp các cung $e_1, e_2, ..., e_n$ sao cho đỉnh cuối của cung e_i là đỉnh đầu của cung e_{i+1} với $1 \leqslant i \leqslant n-1$.

Nhãn λ của đường Π trong đồ hình tổng quát D được xác định bởi

$$h_E(e_1e_2...e_n) = w_1w_2...w_n.$$

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

Đường Π trong đồ hình tổng quát D được biểu diễn như sau:



trong đó:

$w_1 = \psi_1\varphi_1^- = \psi_1\varphi_0^-$,

$w_i = \psi_i\varphi_i^-$ với $2 \leqslant i \leqslant n$,

$\mu_0 = w_0$,

$\mu_i = f(\mu_{i-1}, w_i) = f(\mu_{i-1}, \psi_i\varphi_i^-)$ với $1 \leqslant i \leqslant n$,

$h_V(v_0) = h^{-1}(\omega_0^-) = w_0 \in g(\mu_0)$,
$h_V(v_i) = h^{-1}(\varphi_{i+1}^-) = \varphi_{i+1} \in g(\mu_i)$ với $1 \leqslant i \leqslant n-1$,
$h_V(v_n) = \varphi_{n+1} \in g(\mu_n)$,
và được viết bởi $(v_0, \mu_0)\lambda(v_n, \mu_n)$ với nhãn $\lambda = w_1w_2...w_n$.
Khi đó $\mu_i = f(\mu_{i-1}, w_i)$ là phần ghi nhớ với $1 \leqslant i \leqslant n-1$ có thể được xác định như sau:
$\mu_i = \emptyset^i(\omega_0, w_1, ..., w_i)$ với $1 \leqslant i \leqslant n-1, n$ là độ dài của đường Π trong đồ hình tổng quát $D, \emptyset^n$ là hàm được xác định trên tích Đêcac $(U \cup U^-)^* \times W \times ... \times W$ (n lần) và thỏa mãn các điều kiện sau:

$$\emptyset^0(\omega) = \omega \quad \text{với } \omega \in (U \cup U^-)^*,$$

$$\emptyset^i(\omega, w_1, ..., w_i) = f(\emptyset^{i-1}(\omega, w_1, ..., w_{i-1}), w_i) \quad \text{với } \omega \in (U \cup U^-)^*, w_i \in W,\ 1 \leqslant i \leqslant n.$$

Đường trong đồ hình tổng quát D được gọi là đường đơn nếu đường đó đi qua mỗi cung không quá một lần.

Đường trong đồ hình tổng quát D được gọi là chu trình nếu đường đó bắt đầu và kết thúc tại cùng một đỉnh.

Chu trình trong đồ hình tổng quát D được gọi là chu trình đơn nếu chu trình đó đi qua mỗi cung không quá một lần.

Đường Π trong đồ hình tổng quát D được gọi là đường đầy đủ với nhãn λ nếu đường đó đi từ đỉnh ban đầu v_0 với phần ghi nhớ ban đầu μ_0 tới đỉnh kết thúc v_f với phần ghi nhớ kết thúc μ_f và được viết như sau:

$$(v_0, \mu_0)\lambda(v_f, \mu_f),$$

trong đó:
$\mu_0 = \sigma_D$,
$\mu_i = \emptyset^i(\sigma_D, w_1, ..., w_i)$ với $1 \leqslant i \leqslant n-1$,
$\mu_f = \emptyset^n(\sigma_D, w_1, ..., w_n)$.

Dãy $M_D = (\mu_0, \mu_1, ..., \mu_i)$ được gọi là dãy ghi nhớ trong đồ hình tổng quát D với độ dài $|M_D| = i,\ 1 \leqslant i \leqslant n-1$.

Dãy ghi nhớ M_D được gọi là dãy ghi nhớ kết thúc nếu phần ghi nhớ kết thúc, ký hiệu bởi μ_f, thỏa mãn các điều kiện sau:

$$h_N(\mu_f) = \varepsilon,$$

$$h_T(\mu_f) \in \Sigma_T^*.$$

Dãy ghi nhớ M_D được gọi là dãy ghi nhớ đầy đủ nếu $\mu_0 = \sigma_D, h_N(\mu_f) = \varepsilon$ và $h_T(\mu_f) \in \Sigma_T^*$.

Khi đó $M_D = (\mu_0, \mu_1, ..., \mu_f)$ là dãy ghi nhớ đầy đủ của đường đầy đủ Π trong đồ hình tổng quát D với phần ghi nhớ kết thúc μ_f có tính chất sau:

$$h_N(\mu_f) = \varepsilon,$$

$$h_T(\mu_f) = \omega \in \Sigma_T^*.$$

Dãy $M_D = (\mu_0, \mu_1, ..., \mu_n)$ được gọi là dãy ghi nhớ rút gọn trong đồ hình tổng quát D nếu μ_i là phần ghi nhớ rút gọn với $0 \leqslant i \leqslant n$.

Dãy ghi nhớ M_D được gọi là dãy ghi nhớ kết thúc rút gọn nếu M_D là dãy ghi nhớ rút gọn với phần ghi nhớ kết thúc $\mu_f \in \Sigma_T^*$.

Dãy ghi nhớ M_D được gọi là dãy ghi nhớ đầy đủ rút gọn nếu M_D là dãy ghi nhớ rút gọn với $\mu_0 = \sigma_D$ và $\mu_f \in \Sigma_T^*$.

Khi đó $M_D = (\mu_0, \mu_1, ..., \mu_f)$ là dãy ghi nhớ đầy đủ rút gọn của đường đầy đủ Π trong đồ hình tổng quát D có tính chất sau:

$$\mu_i \in \Sigma_N \cup \Sigma_T)^*,$$

$$\mu_f = \omega \in \Sigma_T^*.$$

Như vậy ω là xâu được xác định bởi đường đầy đủ Π trong đồ hình tổng quát D.

Nhận xét 1

- Do $\Sigma_N, \Sigma_N^-, \Sigma_T, \Sigma_T^-, U_V, W$ là các tập hữu hạn nên tập các đỉnh V và tập các cung E cũng là hữu hạn. Vì vậy đồ hình tổng quát D được xem như là đồ thị hữu hạn có hướng được gán nhãn và tại các đỉnh có các phần ghi nhớ.

- Đường đầy đủ Π trong đồ hình tổng quát D với nhãn λ được cấu thành từ đường đầy đủ đơn và các chu trình đơn (nếu có).

- Phần ghi nhớ kết thúc μ_f trong dãy ghi nhớ đầy đủ rút gọn M_D là xâu gồm các ký hiệu kết thúc và xem như là xâu được xác định bởi đường đầy đủ trong đồ hình tổng quát D.

Định nghĩa 6. Ngôn ngữ được xác định bởi đồ hình tổng quát D là tập:

$$L(D) = \{\omega \in \Sigma_T^* | \omega = h_T(\mu_f), h_N(\mu_f) = \varepsilon, (v_0, \mu_0)\lambda(v_f, \mu_f)\}.$$

4. SỰ TƯƠNG ĐƯƠNG GIỮA ĐỒ HÌNH TỔNG QUÁT VÀ VĂN PHẠM CẤU TRÚC - CÂU

Bổ đề 1. *Đối với đồ hình tổng quát D luôn tồn tại văn phạm cấu trúc - câu G, sao cho $L(G) = L(D)$.*

Chứng minh. Giả sử $D = (\Sigma_N, \Sigma_N^-, \Sigma_T, \Sigma_T^-, W, V, E, H, v_0, v_f)$ là đồ hình tổng quát.

Văn phạm cấu trúc - câu $G = (N, \Sigma, P, \sigma)$ được xác định như sau:

$N = \Sigma_N$,

$\Sigma = \Sigma_T$,

$\sigma = \sigma_D, \sigma_D$ là nhãn của đỉnh ban đầu v_0,

P là tập các quy tắc có dạng sau:

$$p = h^{-1}(\varphi^-) \to \psi = \varphi \to \psi \text{ với } w = \psi\varphi^- \in W.$$

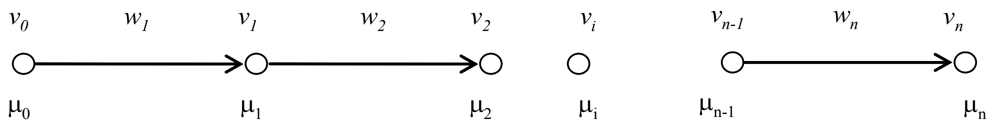
Ta cần chứng minh $L(G) = L(D)$.

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

Xét xâu bất kỳ khác rỗng $\omega \in L(D)$.

Do $\omega \in L(D)$ nên tồn tại đường đầy đủ Π trong đồ hình tổng quát D với nhãn $\lambda = w_1 w_2 ... w_n$ và dãy ghi nhớ đầy đủ rút gọn $M_D = (\mu_0, \mu_1, ..., \mu_f)$, sao cho $\mu_f = \omega$.

Đường đầy đủ Π trong đồ hình tổng quát D được biểu diễn như sau:



trong đó:

$w_1 = \psi_1\varphi_1^- = \psi_1\sigma_D^-$,
$w_i = \psi_i\varphi_i^-$ với $2 \leqslant i \leqslant n$,
$\mu_0 = \emptyset^0(\sigma_D) = \sigma_D$,
$\mu_i = \emptyset^i(\sigma_D, w_1, ..., w_i)$ với $1 \leqslant i \leqslant n-1$,
$\mu_f = \emptyset^n(\sigma_D, \lambda)$,
$h_V(v_0) = h^{-1}(\sigma_D^-) = \sigma_D \in g(\mu_0)$,
$h_V(v_i) = h^{-1}(\varphi_{i+1}^-) = \varphi_{i+1} \in g(\mu_i)$ với $1 \leqslant i \leqslant n-1$,
$h_V(v_f) = K$ và thỏa mãn điều kiện $K \to \varepsilon$.

Khi đó ta có:

$\sigma = \sigma_D$,
$p_1 = h^{-1}(\sigma^-) \to \psi_1 = \sigma \to \psi_1$,
$p_i = h^{-1}(\varphi_i^-) \to \psi_i = \varphi_i \to \psi_i$ với $2 \leqslant i \leqslant n$,
$\omega_0 = \mu_0 = \sigma_D = \sigma$,
$\omega_i = \mu_i$ với $1 \leqslant i \leqslant n-1$,
$\omega_n = \mu_f$,
và dẫn xuất đầy đủ $B = (\omega_0, \omega_1, ..., \omega_n)$ trong văn phạm cấu trúc - câu G với xâu kết thúc $\omega_n = \mu_f = \omega$. Suy ra $\omega \in L(G)$. Vậy $L(D) \subseteq L(G)$.

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

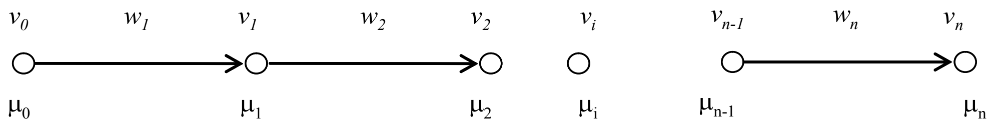
Xét xâu bất kỳ khác rỗng $\omega \in L(G)$. Do $\omega \in L(G)$ nên tồn tại dãy các quy tắc $p_1, p_2, ..., p_n$ và dẫn xuất đầy đủ $B = (\omega_0, \omega_1, ..., \omega_n)$ trong văn phạm cấu trúc - câu G sao cho $\omega_n = \omega$, trong đó:

$\omega_0 = \sigma$,
$p_1 = \sigma \to \psi_1$,
$p_i = \varphi_i \to \psi_i$ với $2 \leqslant i \leqslant n$.

Khi đó ta có:

$\sigma_D = \sigma$,
$w_1 = \psi_1 h(\sigma_D) = \psi_1\sigma_D^-$,
$w_i = \psi_1 h(\varphi_i) = \psi_1\varphi_i^-$ với $2 \leqslant i \leqslant n$,
$\mu_0 = \omega_0 = \sigma = \sigma_D$,
$\mu_i = \omega_i$ với $1 \leqslant i \leqslant n-1$,
$\mu_f = \omega_n$,
$v_0 = h^{-1}V(\sigma_D)$ với $\sigma_D \in g(\mu_0)$,
$v_i = h^{-1}V(\varphi_{i+1})$ với $\varphi_{i+1} \in g(\mu_i), 1 \leqslant i \leqslant n-1$,
$v_f = h^{-1}V(K)$ và thỏa mãn điều kiện $K \to \varepsilon$,

và đường đầy đủ Π trong đồ hình tổng quát D với nhãn $\lambda = w_1w_2...w_n$ được biểu diễn như sau:



Suy ra dãy ghi nhớ đầy đủ rút gọn $M_D = (\mu_0, \mu_1, ..., \mu_f)$ của đường đầy đủ Π với phần ghi nhớ kết thúc $\mu_f = \omega_n = \omega \in L(D)$. Vậy $L(G) \subseteq L(D)$. Kết luận $L(G) = L(D)$. Bổ đề đã được chứng minh. ■

Bổ đề 2. *Đối với văn phạm cấu trúc - câu G luôn tồn tại đồ hình tổng quát D, sao cho* $L(D) = L(G)$.

Chứng minh. Chứng minh $L(D) = L(G)$ tương tự như ở bổ đề trên. Từ Bổ đề 1 và Bổ đề 2 ta có định lý sau: ■

Định lý 6. *Lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát trùng với lớp các ngôn ngữ được sinh bởi các văn phạm cấu trúc - câu.*

Nhận xét 2

- Nhãn của đỉnh ban đầu trong đồ hình tổng quát là ký hiệu ban đầu của văn phạm cấu trúc - câu và nhãn của các đỉnh trong đồ hình tổng quát chính là vế trái của các quy tắc trong văn phạm cấu trúc - câu.
- Dãy ghi nhớ đầy đủ rút gọn trong đồ hình tổng quát chính là dẫn xuất đầy đủ trong văn phạm cấu trúc - câu.
- Tập các đường đầy đủ trong đồ hình tổng quát và tập các dẫn xuất đầy đủ trong văn phạm cấu trúc - câu tương ứng có sự tương ứng 1–1.

5. SỰ PHÂN LOẠI ĐỒ HÌNH TỔNG QUÁT

Trên cơ sở định nghĩa đồ hình tổng quát và dựa trên đặc điểm của tập các xâu ghép tiếp ta có thể phân đồ hình tổng quát thành bốn loại:

Đồ hình tổng quát loại 0

Đồ hình tổng quát D được gọi là đồ hình tổng quát loại 0, nếu mỗi xâu ghép tiếp có dạng $\psi\varphi^-$ và không có sự hạn chế, ràng buộc nào.

Lớp các đồ hình tổng quát loại 0, được ký hiệu bởi $\mathcal{D}_o$ là tập các đồ hình tổng quát loại 0.

Lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát loại 0, được ký hiệu bởi $\mathcal{L}(\mathcal{D}_0)$, là tập các ngôn ngữ được xác định bởi các đồ hình tổng quát loại 0.

Đồ hình tổng quát loại 1

Đồ hình tổng quát D được gọi là đồ hình tổng quát loại 1, nếu mỗi xâu ghép tiếp có dạng $\psi\varphi^-$ với $|\varphi^-| \leqslant |\psi|, \varphi^- \in (U^-)^*, \psi \in U^*$.

Lớp các đồ hình tổng quát loại 1, được ký hiệu bởi $\mathcal{D}_1$ là tập các đồ hình tổng quát loại 1.

Lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát loại 1, ký hiệu bởi $\mathcal{L}(\mathcal{D}_1)$ là tập các ngôn ngữ được xác định bởi các đồ hình tổng quát loại 1.

Đồ hình tổng quát loại 2

Đồ hình tổng quát D được gọi là đồ hình tổng quát loại 2, nếu mỗi xâu ghép tiếp có dạng αA^- với $A^- \in \Sigma_N^-$ và $\alpha \in U^*$.

Lớp các đồ hình tổng quát loại 2, được ký hiệu bởi $\mathcal{D}_2$ là tập các đồ hình tổng quát loại 2.

Lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát loại 2, ký hiệu bởi $\mathcal{L}(\mathcal{D}_2)$ là tập các ngôn ngữ được xác định bởi các đồ hình tổng quát loại 2.

Đồ hình tổng quát loại 3

Đồ hình tổng quát được gọi là đồ hình tổng quát loại 3, nếu mỗi xâu ghép tiếp có một trong hai dạng sau:

i) aBA^-,

ii) aA^-,

với $A^- \in \Sigma_N^-, B \in \Sigma_N$ và $a \in \Sigma_T$.

Lớp các đồ hình tổng quát loại 3, được ký hiệu bởi $\mathcal{D}_3$ là tập các đồ hình tổng quát loại 3.

Lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát loại 3, được ký hiệu bởi $\mathcal{L}(\mathcal{D}_3)$ là tập các ngôn ngữ được xác định bởi các đồ hình tổng quát loại 3.

Lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát tương ứng với sự phân loại trên thỏa mãn bao hàm thức sau:

$$\mathcal{L}(\mathcal{D}_3) \subset \mathcal{L}(\mathcal{D}_2) \subset \mathcal{L}(D_1) \subset \mathcal{L}(D_0).$$

Dựa trên định nghĩa văn phạm cấu trúc - câu, đồ hình tổng quát và từ sự phân loại văn phạm cấu trúc - câu, sự phân loại đồ hình tổng quát, ta có nhận xét sau:

Nhận xét 3

Đặc điểm của tập các xâu ghép tiếp trong các đồ hình tổng quát loại 0, đồ hình tổng quát loại 1, đồ hình tổng quát loại 2, đồ hình tổng quát loại 3 giống đặc diểm của tập các quy tắc trong các văn phạm loại 0, văn phạm loại 1, văn phạm loại 2, văn phạm loại 3 và sự phân loại đồ hình tổng quát hoàn toàn giống sự phân loại văn phạm được đưa ra bởi N. Chomsky.

Từ Nhận xét 3 và Định lý 6 hiển nhiên ta có định lý sau:

Định lý 7. *Lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát loại i trùng với lớp các ngôn ngữ được sinh bởi các văn phạm loại i với* $i = 0, 1, 2, 3$.

Định lý 8. *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 ôtômat loại i;*

(3) $\mathcal{L}_i = \mathcal{L}(\mathcal{D}_i)$ *là lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát loại i với* $i = 0, 1, 2, 3$.

Chứng minh

(i) Từ Định lý 5 ta có: (2) được suy ra từ (1) và ngược lại (1) được suy ra từ (2).

(ii) Từ Định lý 7 ta có: (3) được suy ra từ (1) và ngược lại (1) được suy ra từ (3).
(iii) Kết hợp (ii) với (i) ta có: (3) được suy ra từ (2) và ngược lại (2) được suy ra từ (3).

Định lý được chứng minh. ■

6. KẾT LUẬN

Lớp các ngôn ngữ được xác định bởi đồ hình tổng quát trùng với lớp các ngôn ngữ được sinh bởi văn phạm cấu trúc - câu. Đặc biệt lớp các ngôn ngữ được sinh bởi các văn phạm loại i, lớp các ngôn ngữ được đoán nhận bởi các ôtômat loại i và lớp các ngôn ngữ được xác định bởi các đồ hình tổng quát 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), đồ hình tổng quát làm phong phú thêm lý thuyết ngôn ngữ hình thức, ôtômat và ứng dụng 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] Arto Salomaa, *Nhập môn tin học lý thuyết tinh toán các ôtômat,* NXB Khoa học và kỹ thuật, Hà Nội, 1992.

[2] Nguyễn Văn Ba, *Ngôn ngữ hình thức*, NXB Khoa học và kỹ thuật, Hà Nội, 2002.

[3] Đặng Huy Ruận, Nguyễn Thế Bình, Nguồn và độ phức tạp tính toán của nguồn, *Kỷ yếu Hội thảo Quốc gia 2003 (lần thứ 6), Một số vấn đề chọn lọc của công nghệ thông tin* (báo cáo toàn văn), NXB Khoa học và kỹ thuật, Hà Nội, 2005, 311–320.

[4] Đặng Huy Ruận, Nguyễn Thế Bình, Sự mở rộng ngôn ngữ hình thức, *Kỷ yếu Hội thảo Quốc gia 2004 (lần thứ 7), Một số vấn đề chọn lọc của công nghệ thông tin* (báo cáo toàn văn) NXB Khoa học và kỹ thuật, Hà Nội, 2005, 334–344.

[5] Đỗ Đức Giáo, *Cơ sở toán trong lập trình*, NXB Khoa học và kỹ thuật, Hà Nội, 1998.

[6] Đỗ Đức Giáo, Đặng Huy Ruận, *Văn phạm và ngôn ngữ hình thức*, NXB Khoa học và kỹ thuật, Hà Nội, 1991.

[7] Phùng Văn Ổn, "Nghiên cứu một số lớp siêu ngôn ngữ", Luận án Tiến sĩ Toán học, Đại học Khoa học tự nhiên - Đại học Quốc gia Hà Nội, 2001.

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

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

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

Nhận bài ngày 22 - 3 - 2005

Nhận lại sau sửa ngày 22 - 11 - 2005