Tạp chí Tin học và Điều khiển học, T.16, S.2 (2000), 5-8

MỘT SỐ TÍNH CHẤT CỦA LỚP SIÊU NGÔN NGỮ PHI NGỮ CẢNH

ĐẶNG HUY RUẬN, PHÙNG VĂN ỔN

Abstract. In this paper we show some properties of context free languages with infinitive words (we call context free hyper language) and its relation with pushdown automata.

1. MỞ ĐẦU

Có nhiều công trình của nhiều tác giả nghiên cứu về lớp ngôn ngữ chính quy với từ vô hạn (siêu ngôn ngữ chính quy) đã được công bố [1-3]). Một cách tự nhiên nảy sinh vấn đề nghiên cứu các tính chất lớp ngôn ngữ phi ngữ cảnh với từ vô hạn (được gọi là siêu ngôn ngữ phi ngữ cảnh). Cũng đã có một số kết quả nghiên cứu về lớp ngôn ngữ này [4,5]. Bài báo này trình bày một số kết quả về vấn đề nêu trên.

2. MỘT SỐ KIẾN THỨC CHUẨN BỊ

2.1. Siêu ngôn ngữ

Cho Σ là tập hữu hạn, không rỗng các chữ cái. Ký hiệu Σ^∞ (Σ^*) là tập hợp tất cả các dãy vô hàn (hữu hạn) các chữ cái trong Σ. Các phần tử của Σ^∞ (Σ^*) được gọi là siêu từ (từ hay từ hữu hạn). Với từ $P \in \Sigma^*$ ký hiệu $|P|$ là độ dài của P, tức là số chữ cái xuất hiện trong P.

Cho $L \subseteq \Sigma^*$, ta định nghĩa hai siêu ngôn ngữ sau:

$$L^\infty = \{P \in \Sigma^\infty \mid P = P_1P_2...P_i... \text{ mà } P_i \in L \text{ và } |P_i| \geq 1 \text{ với mọi } i \geq 1\},$$

$$\lim L = \{P \in \Sigma^\infty \mid P = P_1P_2...P_i... \text{ mà } P_1, P_2, ..., P_i \in L \text{ với mọi } i \geq 1\}.$$

Siêu từ $P \in \Sigma^*$ được gọi là tận cùng tuần hoàn nếu $P = P_1P_2^\infty$ với $P_1, P_2 \in \Sigma^*$.

2.2. Siêu ngôn ngữ phi ngữ cảnh

Cho văn phạm phi ngữ cảnh $G = (\Sigma, V, X, \mathcal{P})$, trong đó $\Sigma = \{a, b\}$, $V = \{X, Y\}$, $\mathcal{P} = \{X \to YX, Y \to aYb \mid ab\}$.

Xét dẫn xuất trái nhất vô hạn:

$$X \to YX \to abX \to abYX \to ababX \to ... \tag{1}$$

ta nhận được siêu từ $(ab)^\infty$. Nếu dùng dẫn xuất trái nhất vô hạn mà sử dụng vô hạn lần biến Y với quy tắc $Y \to aYb$ trong mỗi lần thay thế:

$$X \to YX \to abX \to abYX \to abaYbX \to abaaYbbX \to ... \tag{2}$$

ta nhận được siêu từ $ab(a)^\infty$. Trong trường hợp này, phần vô hạn ở bên phải được bỏ qua.

Như vậy có hai cách sinh ra siêu ngôn ngữ phi ngữ cảnh, tùy thuộc vào việc sử dụng dẫn xuất trái nhất (1) hoặc các biến được dùng vô hạn lần được xác định (2).

Định nghĩa 1. Cho văn phạm phi ngữ cảnh $G = (\Sigma, V, X, \mathcal{P})$.

a) Siêu ngôn ngữ $L \subseteq \Sigma^\infty$ được gọi là *đại số* nếu có một văn phạm phi ngữ cảnh G thỏa mãn L bao gồm tất cả các siêu từ của Σ^∞, mà chúng được sinh từ G bởi dẫn xuất trái nhất.

b) Siêu ngôn ngữ $L \subseteq \Sigma^\infty$ được gọi là *phi ngữ cảnh* nếu có một văn phạm phi ngữ cảnh và tập $\mathcal{F} = 2^V$ (kỳ hiệu 2^V là tập tất cả các tập con của tập V, kể cả tập rỗng) sao cho L bao gồm tất cả các từ Σ^∞ mà chúng được sinh từ G bởi dẫn xuất trái nhất, trong đó các biến được sử dụng vô hạn lần tạo nên một tập trong $\mathcal{F}$.

Theo [5], lớp siêu ngôn ngữ đại số hẹp hơn lớp siêu ngôn ngữ phi ngữ cảnh.

Giả sử G là văn phạm phi ngữ cảnh. Ký hiệu $\mathcal{CF}$ và $\mathcal{DCF}$ là họ các ngôn ngữ phi ngữ cảnh và phi ngữ cảnh đơn định. Ta có $\mathcal{DCF} \subseteq \mathcal{CF}$.

Gọi $\mathcal{L}$ là một lớp các ngôn ngữ từ hữu hạn nào đó ($\subseteq A^*$). Bao đóng ∞ - Kleene của $\mathcal{L}$ là lớp tất cả các siêu ngôn ngữ và là hợp hữu hạn của các tập UV^∞ với $U, V \in \mathcal{L}$. Trong [5] đã chỉ ra mối quan hệ của lớp siêu ngôn ngữ phi ngữ cảnh với ngôn ngữ phi ngữ cảnh từ hữu hạn:

Định lý 1. [5] *Siêu ngôn ngữ là phi ngữ cảnh khi và chỉ khi nó thuộc bao đóng ∞ - Kleene của lớp các ngôn ngữ phi ngữ cảnh.*

2.3. Siêu otomat đẩy xuống

Siêu otomat đẩy xuống là một otomat đẩy xuống đoán nhận một siêu từ mà hoạt động của nó được mô tả như sau:

Cho otomat đẩy xuống $M = (S, \Sigma, V, f, s_0, Z_0)$ trong đó S là tập hữu hạn, không rỗng các trạng thái; Σ là bảng chữ cái vào; V là bảng chữ cái của ngăn xếp; $s_0 \in S$ là trạng thái khởi đầu; $Z_0 \in V$ là chữ cái đặc biệt dùng làm đáy của ngăn xếp; f là hàm bộ phận từ $S \times (\Sigma \cup \{\varepsilon\}) \times V$ vào tập các tập con hữu hạn của $S \times V^*$, còn gọi là hàm chuyển.

Cặp $C = (s, Q)$ được gọi là hình trạng của otomat, biểu thị trạng thái hiện tại s và xâu Q hiện có trong ngăn xếp của otomat.

Một bước chuyển của M, ký hiệu là $(s, QZ) \xrightarrow{a} (s', QQ_1)$ nếu $(s', Q_1) \in f(s, a, Z)$ trong đó $a \in (\Sigma \cup \{\varepsilon\})$, $s, s' \in S$ và otomat sẽ chuyển từ trạng thái s sang trạng thái s'; $Q, Q_1 \in V^*$; $Z \in V$ là ký hiệu trên đỉnh của ngăn xếp và nó sẽ được thay thế bởi xâu Q_1. Trong trường hợp $a = \varepsilon$, ta gọi là ε - bước chuyển.

Một dãy hữu hạn các bước chuyển $C_0 \xrightarrow{a_1} C_1 \xrightarrow{a_2} C_2 \ldots \xrightarrow{a_n} C_n$ trong đó $n > 0$, $a_1 \in (\Sigma \cup \{\varepsilon\})$ với $1 \le i \le n$; $C_i = (s_i, Q_i)$ với $0 \le i \le n$ được gọi là C - tính toán (ký hiệu $P = a_1a_2...a_n$ gọi là nhãn của C - tính toán) và được viết $C = C_0C_1...C_n$ nếu nhãn P là rõ ràng hoặc viết ngắn gọn $C_0 \xrightarrow{P} *C_n$.

Tương tự, một dãy vô hạn các bước chuyển của otomat được gọi là ∞ - tính toán và ký hiệu là $C_\infty = C_0C_1...$ Với ∞ - tính toán $C_\infty = C_0C_1...$ ta ký hiệu:

$$\mathrm{In}(C_n) = \{s \in S \mid s = s_n \text{ với vô hạn các chỉ số } n, \text{ trong đó } C_n = (s_n, Q_n)\}$$

Giả sử $F \subseteq S$ và $\mathcal{F} \subseteq 2^s$. Một ∞ - tính toán $C_\infty = C_0C_1...$ gọi là được chấp nhận nếu $\mathrm{In}(C_\infty) \cap F \ne \emptyset$ trong đó $C_0 = (s_0, Z_0)$ và $a_n \ne \varepsilon$ với vô hạn các chỉ số n. Tương tự, ∞ - tính toán $C_\infty = C_0C_1...$ gọi là *được chấp nhận ngặt* nếu $\mathrm{In}(C_\infty) \in \mathcal{F}$.

Định nghĩa 2. Siêu ngôn ngữ được đoán nhận bởi siêu otomat đẩy xuống M xác định như sau:

$L_\infty(M, F) = \{P \in \Sigma^\infty \mid \mathrm{In}(C_\infty) \cap F \ne \emptyset$ với P là nhãn của một lớp ∞ - tính toán $C_\infty\}$,

$L'_\infty(M, \mathcal{F}) = \{P \in \Sigma^\infty \mid \mathrm{In}(C_\infty) \in \mathcal{F}$ với P là nhãn của một lớp ∞ - tính toán $C_\infty\}$.

Ký hiệu $\mathcal{P}_\infty$, $\mathcal{P}'_\infty$ là 2 họ các siêu ngôn ngữ được đoán nhận bởi siêu otomat đẩy xuống tương ứng với 2 cách đoán nhận trên. Theo [4] ta có:

Định lý 2. $\mathcal{P}_\infty = \mathcal{P}'_\infty$.

Cũng theo [4] ta có:

Định lý 3. *Nếu $L \in \mathcal{P}_\infty$ và $L \ne \emptyset$ thì L chứa ít nhất một siêu từ tận cùng tuần hoàn.*

Định nghĩa 3. Siêu otomat đẩy xuống M được gọi là đơn định nếu và chỉ nếu hàm chuyển $f : S \times (\Sigma \cup \{\varepsilon\} \times V \to S \times V^*$ thỏa mãn $f(s, a, Z)$ có không quá một phần tử và với mỗi $s \in S$, $Z \in V$, nếu $f(s, \varepsilon, Z)$ là xác định thì $f(s, a, Z)$ không xác định với mọi $a \in \Sigma$.

Ký hiệu $\mathcal{DP}_\infty$, $\mathcal{DP}'_\infty$ là 2 họ các siêu ngôn ngữ được đoán nhận bởi siêu otomat đẩy xuống tương ứng với 2 cách đoán nhận nêu trong Định nghĩa 3. Theo [4] ta có:

Định lý 4. $\mathcal{DP}_\infty \subseteq \mathcal{DP}'_\infty \subseteq \mathcal{P}_\infty \subseteq \mathcal{P}'_\infty$.

3. CÁC KẾT QUẢ CHÍNH

Định lý 5. *Mỗi siêu ngôn ngữ phi ngữ cảnh (không rỗng) đều chứa ít nhất một siêu từ tận cùng tuần hoàn.*

Chứng minh. Giả sử L là siêu ngôn ngữ phi ngữ cảnh. Theo Định lý 1, thì L sẽ thuộc bao đóng ∞ - Kleene của lớp các ngôn ngữ phi ngữ cảnh, nghĩa là có thể biểu diễn L theo dạng:

$L = \bigcup_{i=1}^{n} U_i V_i^\infty$ với U_i, V_i là những ngôn ngữ phi ngữ cảnh nào đó.

Điều đó có nghĩa là, với mỗi siêu từ $P \in L$, tồn tại các từ $P_1 \in U_i$, $P_2 \in V_i$ với i nào đó sao cho$P = P_1 P_2^\infty$. Vậy P là siêu từ tận cùng tuần hoàn.

Định lý 6. *Trong $\mathcal{P}_\infty \cap \lim \mathcal{CF}$ có chứa siêu ngôn ngữ phi ngữ cảnh mà nó chứa ít nhất một siêu từ tận cùng tuần hoàn.*

Định lý 6 là hệ quả trực tiếp từ 2 bổ đề đã được nêu bởi Linna [4].

Bổ đề 1. [4] *Cho L là ngôn ngữ trên bảng chữ cái V, và $d \notin V$. Khi đó $Ld^\infty \in \lim \mathcal{CF}$ nếu và chỉ nếu $L \in \mathcal{CF}$.*

Bổ đề 2. [4] *Cho L là ngôn ngữ trên bảng chữ cái V, và $d \notin V$. Khi đó $Ld^\infty \in \mathcal{P}_\infty$ nếu và chỉ nếu $L \in \mathcal{CF}$.*

Định lý 7. *Lớp siêu ngôn ngữ phi ngữ cảnh và lớp $\mathcal{P}_\infty$ là trùng nhau.*

Định lý 7 là hệ trực tiếp từ 2 bổ đề sau:

Bổ đề 3. *Với mỗi siêu ngôn ngữ phi ngữ cảnh L, có thể xây dựng được siêu otomat đẩy xuống $M = (S, \Sigma, V, f, s_0, Z_0)$ sao cho $L = L_\infty(M, F)$ với $F \subseteq S$.*

Chứng minh. Giả sử L là siêu ngôn ngữ phi ngữ cảnh, theo Định lý 1 thì $L = \bigcup_{i=1}^{n} U_i V_i^\infty$ với U_i, V_i là những ngôn ngữ phi ngữ cảnh nào đó. Không mất tính tổng quát, ta xét $L = UV^\infty$.

Gọi $G_u = (\Sigma, V_u, X_u, \mathcal{P}_u)$, $G_v = (\Sigma, V_v, X_v, \mathcal{P}_v)$ là những văn phạm phi ngữ cảnh trong đó $V_u \cap V_v = \emptyset$ sinh ra ngôn ngữ U và V.

Dễ nhận thấy văn phạm $G = (\Sigma, V_u \cup V_v \cup \{X_1, X'_1\}, S_1, \mathcal{P})$ trong đó $\{X_1, X'_1\} \cap (V_u \cup V_v) = \emptyset$ và $\mathcal{P} = \mathcal{P}_u \cup \mathcal{P}_v \cup \{X^* \to X_u X_1, X_1 \to X_v X_1\}$ sinh ra siêu ngôn ngữ $L = UV^\infty$ với $\mathcal{F}$ là tập của tất cả các tập con của tập $(V_v \cup \{X'_1\})$.

Ta xây dựng được siêu otomat đẩy xuống $M_1 = (S, \Sigma, V, f, s_0, Z_0)$ như sau:

$S = \{s_0, s_u, s_v, s_F\}$; $V = \Sigma \cup V_u \cup V_v \cup \{X^*, X_1, Z_0, Z_u, Z_v\}$ trong đó $\{X^*, X_1, Z_0, Z_u, Z_v\} \cap (\Sigma \cup V_u \cup V_v)$; $F = \{s_F\}$; hàm chuyển trạng thái như sau:

1) $f(s_0, \varepsilon, Z_0) = (s_u, X^* Z_u X_u)$
2) Với $X \in V_u$, $f(s_u, \varepsilon, X) = \{(s_u, \alpha^R) \mid X \to \alpha \in \mathcal{P}_u$, với α^R là xâu đảo ngược của $\alpha\}$
3) Với $a \in \Sigma$, $f(s_u, a, a) = \{(s_u, \varepsilon)\}$
4) $f(s_u, \varepsilon, Z_u) = (s_v, \varepsilon)$
5) $f(s_v, \varepsilon, X^*) = (s_v, X^* Z_v X_v)$
6) Với $X \in V_v$, $f(s_v, \varepsilon, X) = \{(s_v, \alpha^R) \mid X \to \alpha \in \mathcal{P}_v$, với α^R là xâu đảo ngược của $\alpha\}$
7) Với $a \in \Sigma$, $f(s_v, a, a) = \{(s_v, \varepsilon)\}$
8) $f(s_v, \varepsilon, Z_v) = (s_F, \varepsilon)$
9) $f(s_F, \varepsilon, X^*) = (s_v, X^*)$

Thực tế là hàm chuyển 1) sẽ đưa hình trạng đầu $C_0 = (s_0, Z_0)$ về hình trạng $C_u = (s_u, X_u)$ để bắt đầu quá trình đoán nhận các từ thuộc U với các hàm 2) và 3). Hàm 4) là khi đã đoán nhận xong từ thuộc U, bắt đầu chuyển sang hình trạng $C_v = (s_v, X_v)$ để bắt đầu quá trình đoán nhận các từ thuộc V với các hàm 5), 6), 7). Hàm 8) là khi đoán xong từ thuộc V, sẽ chuyển về trạng thái kết thúc s_F để rồi sau đó hàm 9) bắt đầu trở lại hình trạng $C_v = (s_v, X_v)$, bắt đầu quá trình đoán nhận tiếp theo từ của V.

Dễ kiểm chứng rằng $L = L_\infty(M, F)$ với $F = \{s_F\}$.

Bổ đề 4. *Với mỗi siêu otomat đẩy xuống $M = (S, \Sigma, V, f, s_0, Z_0)$ ta có thể tìm được siêu ngôn ngữ phi ngữ cảnh L sao cho $L = L_\infty(M, F)$ với $F \subseteq S$.*

Chứng minh. Với mỗi siêu otomat đẩy xuống $M = (S, \Sigma, V, f, s_0, Z_0)$, các trạng thái $s_1, s_2, s_3 \in S$ và các chữ cái $Z_1, Z_2 \in V$, ta ký hiệu:

$$L_{s_1, Z_1} = \{P \in \Sigma^* \mid (s_0, Z_0) \xrightarrow{P}{}^*(s_1, QZ_1) \text{ với } Q \text{ (nào đó)} \in V^*\},$$

$$L^{s_1, Z_1}_{s_2, Z_2(s_3)} = \{P \in \Sigma^* \mid \text{tồn tại } P_1, P_2 : P = P_1P_2 \text{ và } (s_1, Z_1) \xrightarrow{p_1}{}^*(s_3, Q_1) \xrightarrow{p_2}{}^* (s_2, Q_2Z_2) \text{ với } Q_1, Q_2 \text{ (nào đó)} \in V^*\}.$$

Mệnh đề sau đã được chứng minh bởi M. Linna [5]:

Mệnh đề. *Cho otomat đẩy xuống $M = (S, \Sigma, V, f, s_0, Z_0)$. Khi đó $P \in L_\infty(M, F)$ nếu và chỉ nếu tồn tại $s_3 \in S$ và $Z_2 \in V$ sao cho $P \in L_{s,Z}\big(L^{s,Z}_{s,Z_{(s_F)}}\big)^\infty \cap \Sigma^\infty$.*

Bổ đề 4 được suy ra trực tiếp từ mệnh đề trên với việc lấy $L = L_{s,Z}\big(L^{s,Z}_{s,Z_{(s_F)}}\big)^\infty \cap \Sigma^\infty$.

4. KẾT LUẬN

Vấn đề nêu ra đã cơ bản được giải quyết. Đã nêu được một số tính chất của lớp siêu ngôn ngữ phi ngữ cảnh và mối quan hệ của nó với lớp siêu ngôn ngữ được đoán nhận bởi siêu otomat đẩy xuống. Tuy nhiên còn nhiều khía cạnh có thể tiếp tục nghiên cứu như mối quan hệ của lớp siêu ngôn ngữ đại số với otomat đẩy xuống, mối quan hệ của lớp ngôn ngữ đoán nhận bởi siêu otomat đẩy xuống đơn định với lớp ngôn ngữ phi ngữ cảnh đơn định.

TÀI LIỆU THAM KHẢO

[1] Đặng Huy Ruận, Phùng Văn Ổn, Some results of regular hyper-language, *Tạp chí Khoa học Đại học Quốc gia Hà Nội* **XV** (1) (1999).

[2] B. Le Saec, V. R. Dare, R. Seromony, Strong recognition of rational ω-languages, *International Conference Mathematical Foundation of Informatics*, 1999.

[3] B. Le Saec, Saturating right congruences, *Theoretical Informatics and Applications* **24** (6) (1990).

[4] M. Linna, On ω-sets associated with context-free languages, *Information and Control* **31** (1976).

[5] W. Thomas, Automata on infinitive words, Formal model and semantics (Handbook of Theorical Computer Science), Vol. B, 1990.

Nhận bài ngày 20 - 8 - 1999

Đặng Huy Ruận, Khoa Toán - Cơ - Tin học,
Trường Đại học Khoa học Tự nhiên - ĐHQG Hà Nội.

Phùng Văn Ổn, Khoa Công nghệ Thông tin,
Trường Đại học Hàng hải Việt Nam.