BG Ch11 BÀI 4. SỰ ĐOÁN NHẬN NGÔN NGỮ

MỞ ĐẦU

Như chúng ta đã thấy ở trên, các ôtômat hữu hạn có thể được dùng như các bộ đoán nhận ngôn ngữ. Vậy các máy này có thể đoán nhận các tập nào? Mặc dù điều này dường như là một bài toán cực kì khó, nhưng các tập có thể được đoán nhận bởi các ôtômat hữu hạn lại có một đặc trưng khá đơn giản. Bài toán này lần đầu tiên đã được nhà toán học Mỹ Stephen Kleene giải quyết vào năm 1956. Ông ta đã chứng minh được rằng tồn tại một ôtômat hữu hạn đoán nhận một tập hợp nếu và chỉ nếu tập đó có thể được xây dựng từ tập rỗng, xâu rỗng và các xâu chỉ chứa một kí hiệu bằng cách ghép, lấy hợp và lấy các bao đóng Kleene theo một trật tự tuỳ ý. Những tập có thể được xây dựng bằng cách như vậy được gọi là các **tập chính quy.**

Các văn phạm chính quy đã được đinh nghĩa trong mục 11.1. Theo thuật ngữ được sử dụng, sẽ không có gì ngạc nhiên khi nói rằng có một quan hệ giữa các tập chính quy được đoán nhận bởi các ôtômat hữu hạn và các văn phạm chính quy. Đặc biệt, một tập là chính quy nếu và chỉ nếu nó được sinh bởi một văn phạm chính quy.

Cuối cùng, có những tập không thể được đoán nhận bởi các ôtômat hữu hạn. Ta sẽ cho một ví dụ về các tập đó đồng thời xem xét một cách ngắn gọn các mô hình tính toán mạnh hơn, như các ôtômat đẩy xuống (pushdown) và các máy Turing ở cuối mục này.

CÁC TẬP CHÍNH QUY

Các tập chính quy là các tập có thể được tạo bằng cách dùng các phép toán ghép, hợp vào bao đóng Kleene theo một trật tự tùy ý xuất phát từ tập rỗng, xâu rỗng và các tập của các xâu chỉ chứa một kí hiệu. Chúng ta sẽ thấy rằng các tập chính quy là các tập có thể được đoán nhận bởi một ôtômat hữu hạn. Để định nghĩa các tập chính quy, trước hết cần phải định nghĩa các biểu thức chính quy.

ĐỊNH NGHĨA 1. Các biểu thức chính quy trên một tập I được định nghĩa một cách đệ quy như sau:

Kí hiệu Ø là một biểu thức chính quy;

Kí hiệu λ là một biểu thức chính quy;

Kí hiệu x là một biểu thức chính quy, với mọi x Є I;

Các kí hiệu (AB), (A ∪ B) và A* là các biểu thức chính quy với mọi A và B là các biểu thức chính quy.

Mỗi biểu thức chính quy biểu diễn một tập được đặc tả bởi các quy tắc sau:

Ø biểu diễn tập rỗng, tức là tập không chứa xâu nào;

λ biểu diễn tập {λ}, là tập chứa xâu rỗng;

x biểu diễn tập {x} chứa xâu chỉ có một kí hiệu;

(**AB**) biểu diễn sự ghép của các tập được biểu diễn bởi **A** và **B**.

A ∪ **B** biểu diễn sự hợp của các tập được biểu diễn bởi **A** và **B**.

A* biểu diễn bao đóng Kleene của tập được biểu diễn bởi **A**.

Các tập được biểu diễn bởi các biểu thức chính quy được gọi là các **tập chính quy**. Từ đây trở đi, các biểu thức chính quy sẽ được dùng để mô tả các tập chính quy, vì vậy khi ta nói tới tập chính quy A là muốn nói tới tập chính quy được biểu diễn bởi biểu thức chính quy A. Ví dụ sau cho thấy các biểu thức chính quy được dùng để đặc tả các tập chính quy như thế nào.

VÍ DỤ 1. Xác định các xâu trong các tập chính quy được đặc tả bởi các biểu thức chính quy sau: 10^*, $(10)^*$, $0 \cup 01$, $0(0 \cup 1)^*$ và $(0^*1)^*$.

Giải: Các tập chính quy được biểu diễn bởi các biểu thức đó được cho trong bảng 1.(độc giả nên kiểm tra lại).

BẢNG 1.	
Biểu thức	**Xâu**
10*	Một số 1 được tiếp theo bởi một số bất kì số 0 (kể cả không có số 0 nào).
(10)*	Một số bất kì các cặp 10 (kể cả xâu rỗng)
0 ∪ 01	Xâu 0 hoặc xâu 01.

0(0 ∪ 1)*	Xâu bất kì bắt đầu bằng 0.
(0*1)*	Xâu bất kì không kết thúc bằng 0.

ĐỊNH LÍ KLEENE

Năm 1956 Kleene đã chứng minh được rằng các tập chính quy là các tập được đoán nhận bởi một ôtômat hữu hạn. Do đó kết quả quan trọng này được gọi là định lí Kleene.

ĐỊNH LÍ 1. ĐỊNH LÍ KLEENE. Một tập là chính quy nếu và chỉ nếu nó được đoán nhận bởi một ôtômat hữu hạn.

Định lí Kleene là một trong những kết quả trung tâm của lí thuyết ô tô mat. Ta sẽ chỉ chứng minh phần chỉ nếu của định lí này, cụ thể chứng minh rằng mọi tập chính quy đều được đoán nhận bởi một ô tô mat hữu hạn. Sự chứng minh phần nếu in nghiêng tức là chứng minh một tập được đoán nhận bởi một ô tô mat hữu hạn là chính quy, xin dành cho bạn đọc như một bài tập.

Chứng minh. Cần nhớ lại rằng một tập chính quy được định nghĩa qua các biểu thức chính quy, mà các biểu thức chính quy lại được định nghĩa một cách đệ quy. Vì vậy, có thể chứng minh được rằng mọi tập chính quy đều được đoán nhận bởi một ô tô mat hữu hạn nếu chúng ta làm được các việc sau:

1. Chứng minh Ø được đoán nhận bởi một ô tô mat hữu hạn.

2. Chứng minh λ được đoán nhận bởi một ô tô mat hữu hạn.

3. Chứng minh {a} được đoán nhận bởi một ô tô mat hữu hạn với mọi a là một kí hiệu trong I.

4. Chứng minh AB được đoán nhận bởi một ô tô mat hữu hạn với mọi A và B đều đã được đoán nhận.

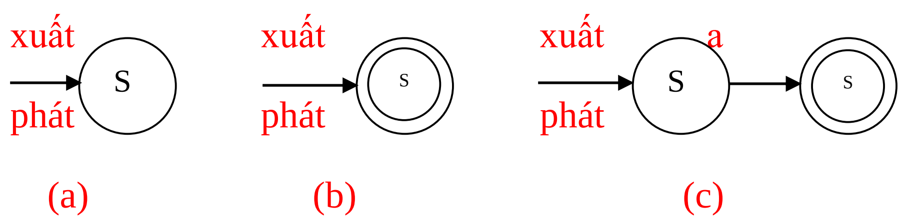
5. Chứng minh A ∪ B được đoán nhận bởi một ô tô mat hữu hạn với mọi A và B đều đã được đoán nhận.

6. Chứng minh A* được đoán nhận bởi một ô tô mat hữu hạn với mọi A đã được đoán nhận.

Bây giờ chúng ta sẽ xét từng nhiệm vụ trên. Trước hết, chứng tỏ rằng Ø được đoán nhận bởi một ôtômat hữu hạn không tất định. Để làm điều này, ta chỉ cần một ôtômat không có các trạng thái kết thúc. Một ôtômat như vậy được cho trên hình 1(a).

Thứ hai, ta sẽ chứng tỏ rằng {λ} được đoán nhận bởi một ôtômat hữu hạn. Để làm điều đó, chỉ cần một ôtômat đoán nhận λ – xâu rỗng, nhưng không đoán nhận một xâu nào khác. Điều này có thể làm được bằng cách làm cho trạng thái xuất phát s_0 là trạng thái kết thúc và không có dịch chuyển nào, vì vậy không có xâu nào khác đưa s_0 đến trạng thái kết thúc. Ôtômat không tất định cho trên hình 1(b) là một máy như vậy.

Thứ 3, ta sẽ chứng tỏ rằng {a} được đoán nhận bởi một ôtômat hữu hạn không tất định. Để làm điều đó, có thể dùng một máy có trạng thái xuất phát s_0 và trạng thái kết thúc s_1. Ta có một chuyển dịch từ s_0 đến s_1 khi đầu vào là 1 và không có một chuyển dịch nào khác. Xâu duy nhất được đoán nhận bởi máy này là a. Máy này được cho trên hình 1(c).



HÌNH 1. Ôtômat hữu hạn không tất định đoán nhận một số tập cơ sở.

Tiếp theo, ta sẽ chứng minh rằng AB và A ∪ B đều có thể được đoán nhận bởi các ô tô mat hữu hạn nếu A và B là các ngôn ngữ đã được đoán nhận bởi các ôtômat hữu hạn. Giả sử A được đoán nhận bởi $M_A = (S_A, I, f_A, s_A, F_A)$.

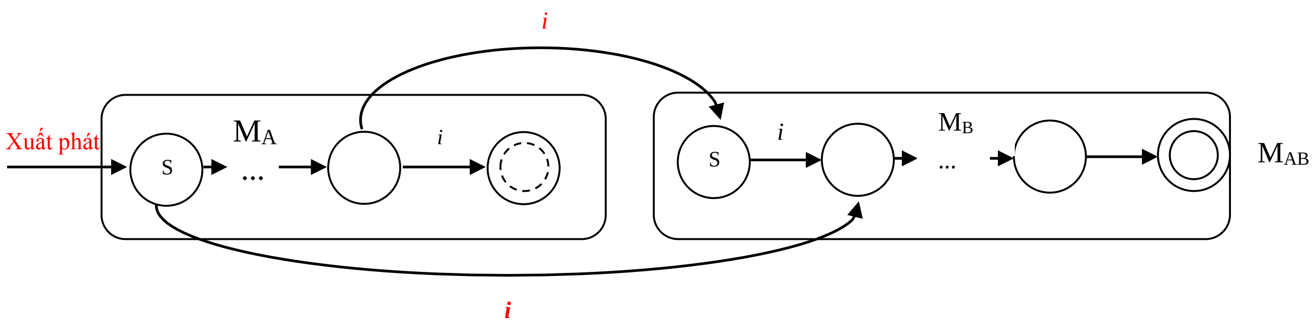
Chúng ta bắt đầu bằng việc dựng máy $M_{AB} = (S_{AB}, I, f_{AB}, s_{AB}, F_{AB})$ đoán nhận AB – ghép của A và B bằng cách tổ hợp các máy cho A và B nối tiếp nhau, sao cho một xâu trong A đưa máy tổ hợp từ s_A – trạng thái xuất phát của M_A đến s_B – trạng thái xuất phát của M_B. Xâu trong B phải đưa máy tổ hợp này từ s_B đến một trạng thái kết thúc của máy tổ hợp. Do đó, ta phải tiến hành chế tạo như sau: Cho $S_{AB} = S_A \cup S_B$. Trạng thái xuất phát s_{AB} như s_A. Tập các trạng thái kết thúc F_{AB} là tập các trạng thái kết thúc của M_B, có chứa cả s_{AB} nếu và chỉ nếu $\lambda \in A \cap B$. Các chuyển dịch trong M_{AB} bao gồm tất cả các chuyển dịch trong M_A và trong M_B, cũng như một số chuyển dịch mới. Đối với mỗi chuyển dịch trong M_A dẫn tới một trạng thái kết thúc, ta sẽ tạo được một chuyển dịch trong M_{AB} từ cùng trạng thái đó tới s_B với cùng một đầu vào. Theo cách đó, một xâu trong A đưa M_{AB} từ s_{AB} đến s_B, rồi sau đó một xâu trong B đưa s_B đến một trạng thái kết thúc của M_{AB}. Hơn nữa, đối với mỗi chuyển dịch từ s_B, chúng ta tạo được một chuyển dịch trong M_{AB} từ s_{AB} tới chính trạng thái đó. Hình 2(a) chứng minh cho sự xây dựng vừa nói ở trên.

Bây giờ chúng ta sẽ dựng máy $M_{A \cup B} = (S_{A \cup B}, I, f_{A \cup B}, s_{A \cup B}, F_{A \cup B})$ đoán nhận $A \cup B$. Ô tô mat này có thể được xây dựn bằng cách tổ hợp M_A và M_B theo kiểu song song, có dùng một trạng thái xuất phát mới với những chuyển dịch mà cả s_A và s_B

đều có. Cho $S_{A \cup B} = S_A \cup S_{B \cup} \{s_{A \cup B}\}$, ở đây $s_{A \cup B}$ là trạng thái mới và cũng là trạng thái xuất phát của $M_{A \cup B}$. Cho tập các trạng thái kết thúc $F_{A \cup B}$ bằng $F_A \cup F_B \cup \{s_{A \cup B}\}$ nếu $\lambda \in A \cup B$ và bằng $F_A \cup F_B$ trong trường hợp còn lại. Các chuyển dịch trong $M_{A \cup B}$ bao gồm tất cả các chuyển dịch trong M_A và trong M_B. Đối với mỗi chuyển dịch từ s_A đến một trạng thái s dưới tác dụng của đầu vào i, ta cũng kể là một chuyển dịch từ $s_{A \cup B}$ đến s cũng dưới tác dụng của đầu vào I và đối với mỗi chuyển dịch từ s_B đến s với đầu vào i ta cũng kể là một chuyển dịch từ $s_{A \cup B}$ đến s với đầu vào i. Theo cách đó, một xâu trong A dẫn từ $s_{A \cup B}$ tới một trạng thái kết thúc trong máy mới và một xâu trong B cũng dẫn từ $s_{A \cup B}$ đến một trạng thái kết thúc trong máy mới. Hình 2(b) minh họa sự xây dựng máy $M_{A \cup B}$ vơaf nói ở trên.

Cuối cùng, chúng ta sẽ xây dựng $M_{A*} = (S_{A*}, I, f_{A*}, s_{A*}, F_{A*})$ – máy đoán nhận A*, tức bao đóng Kleene của A. Cho S_{A*} bao gồm tất cả các trạng thái của A và một trạng thái thêm s_{A*}, là trạng thái xuất phát của máy mới. Tập các trạng thái kết thúc F_{A*} bao gồm tất cả các trạng thái trong F_A cũng như trạng thái xuất phát s_{A*}, vì λ cần phải được đoán nhận. Để đoán nhận ghép của một số tùy ý các xâu thuộc A, chúng ta sẽ đưa vào tất cả các chuyển dịch trong M_A cũng như các chuyển dịch từ s_{A*} khớp với các chuyển dịch từ s_A và các chuyển dịch từ mỗi trạng thái kết thúc khớp với các chuyển dịch từ s_A. Với tập hợp các chuyển dịch đó, một xâu được tạo bởi phép ghép các xâu trong A sẽ đưa s_{A*} tới một trạng thái kết thúc khi xâu thứ hai trong A đã được đọc xong ... Hình 2(c) minh hoạ việc xây dựng mà ta nói ở trên.

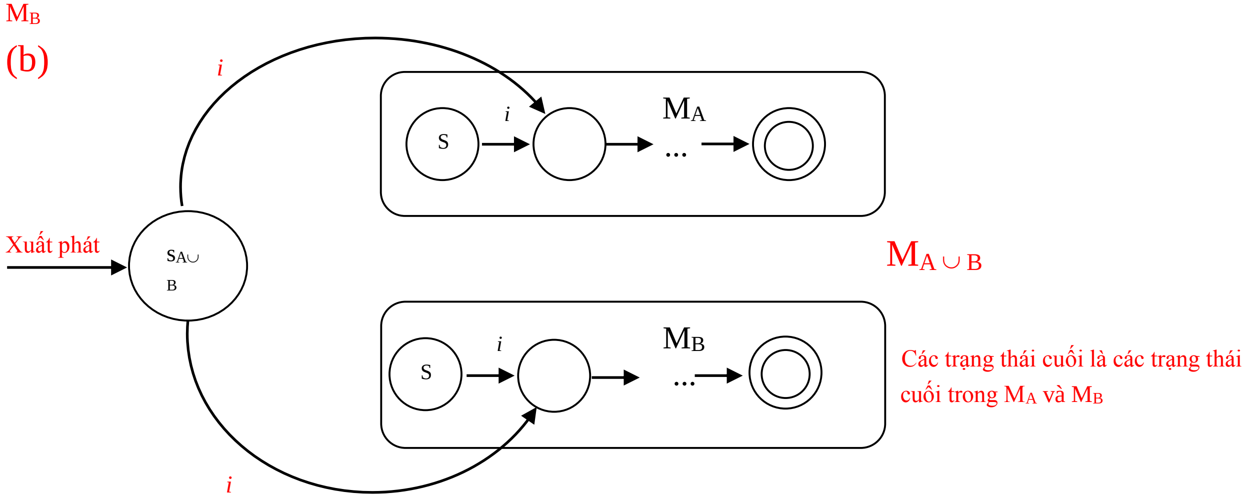
(a) Chuyển trạng thái cuối trong M_A tới s_B



Chuyển từ s_B trong M_B thành $s_{AB} = s_A$

Trạng thái bắt đầu $s_{AB} = s_A$ nếu s_A, s_B là trạng thái cuối   Các trạng thái cuối là tất cả trạng thái kết thúc của M_B

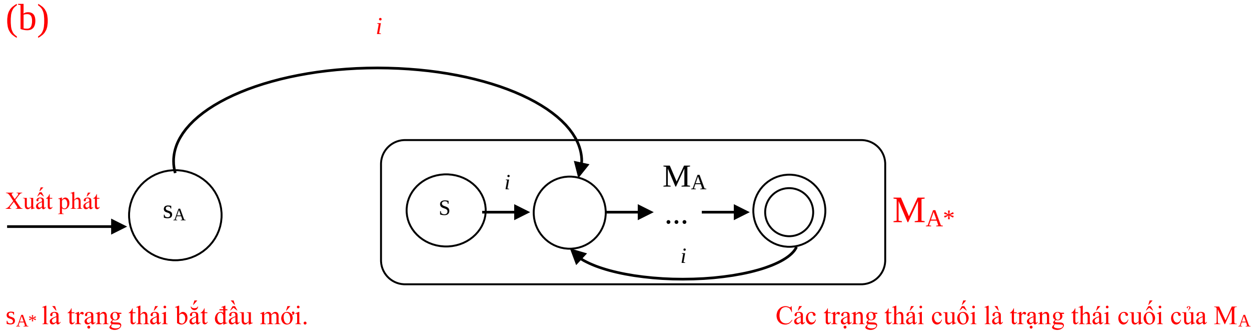
(b)



$s_{A\cup B}$ là trạng thái đầu mới với s_A hoặc s_B là trạng thái cuối.

(c) Chuyển từ s_A thành s_{A*} và tất cả trạng thái kết thúc M_A

(b)



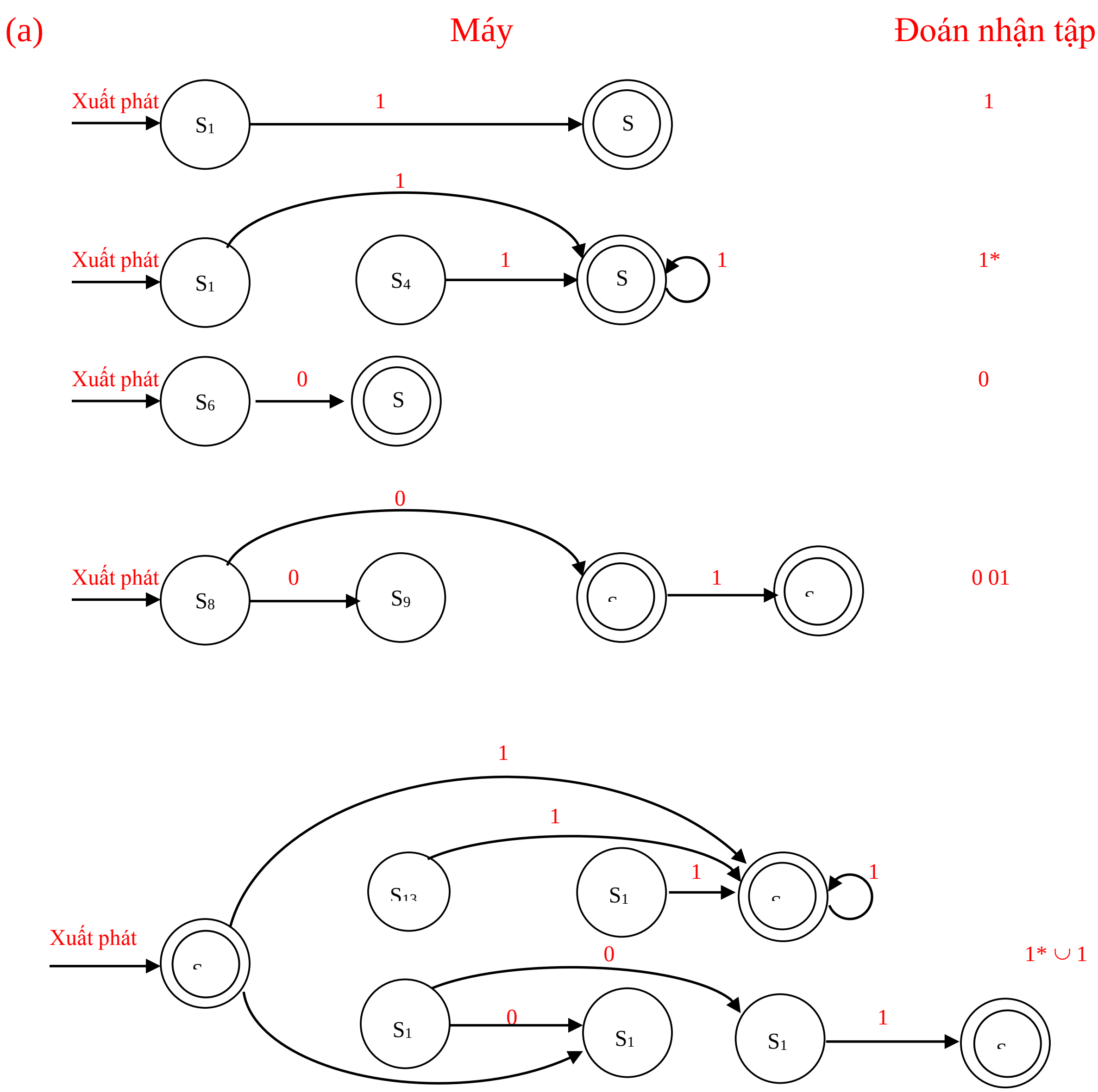
Hình 2: Xây dựng các ôtômat đoán nhận các phép ghép, hợp và bao đóng Kleene.

Có thể xây dựng một ôtômat hữu hạn không tất định cho một tập chính quy bất kì bằng cách dùng thủ tục mô tả trong chứng minh trên. Chúng ta sẽ minh hoạ điều này bằng ví dụ sau.

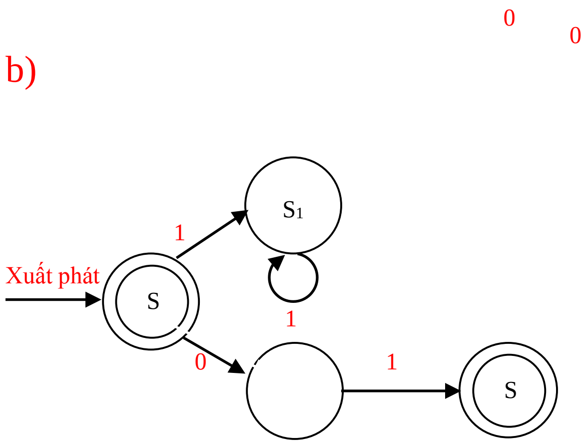
VÍ DỤ 2. Dựng một ôtômat hữu hạn không tất định đoán nhận tập chính quy **1*∪01.**

Giải: Ta bắt đầu bằng việc xây dựng một máy đoán nhận 1*. Điều này được làm bằng cách xây dựng máy đoán nhận 1, rồi sau đó dùng cách dựng cho M_{A*} đã được mô tả trong chứng minh định lí 1. Tiếp theo, dựng máy đoán nhận 01 bằng cách dùng các máy đoán nhận 0 và 1 và cách dựng M_{AB} trong chứng minh định lí 1. Cuối cùng, dùng cách dựng $M_{A\cup B}$ được dùng trong chứng minh đó, chúng ta sẽ dựng được máy đoán nhận **1*∪ 01**. Các ôtômat hữu hạn được dùng để dựng máy này được cho trên hình 3(a).

Những trạng thái trong các máy kế tiếp được đánh dấu bằng cách dùng các chỉ số dưới khác nhau, thậm chí khi một trạng thái được tại thành từ một trạng thái đã được dùng trước đó trong một máy khác. Chú ý rằng cách xây dựng nói trên không tạo ra máy đơn giản nhất đoán nhận **1*∪ 01**. Máy đơn giản hơn đoán nhận tập này được cho trên hình 3(b).



b)



Hình 3: Các ôtômat hữu hạn không tất định đoán nhận $1^* \cup 01$.

TẬP HỢP CHÍNH QUY VÀ VĂN PHẠM CHÍNH QUY

Trong mục 11.1 đã giới thiệu các văn phạm cấu trúc câu và định nghĩa các loại văn phạm khác nhau. Đặc biệt, ta đã định nghĩa văn phạm chính quy hay văn phạm loại 3, là văn phạm có dạng G = (V, T, S, P) với mỗi sản xuất có dạng $S \rightarrow \lambda$, $A \rightarrow a$ hay $A \rightarrow aB$, trong đáo a là ký hiệu kết thúc, A và B là các ký hiệu không kết thúc. Như tên gọi đã gợi ý, giữa các văn phạm chính quy và các tập (hợp) chính quy có một mối quan hệ khăng khít.

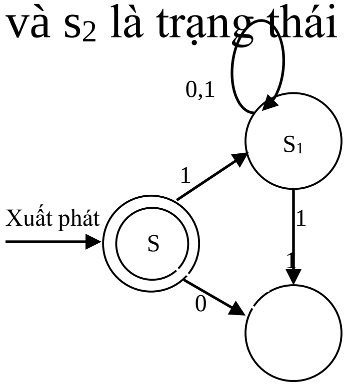
ĐỊNH LÝ 2. Một tập sinh bởi một văn phạm chính quy nếu và chỉ nếu nó là một tập chính quy.

Chứng minh. Trước hết, ta chứng minh rằng một tập được sinh bởi một văn phạm chính quy là một tập chính quy. Giả sử rằng G = (V, T, S, P) là một văn phạm chính quy sinh ra tập L(G). Để chứng minh L(G) là chính quy, ta sẽ xây dựng một máy hữu hạn không tất định M = (S, I, f, s_0, F) đoán nhận L(G). Giả sử S, tập hợp các trạng thái, có chứa trạng thái s_A đối với mỗi kí hiệu không kết thúc A của G và trạng thái phụ của s_F là một trạng thái kết thúc. Trạng thái xuất phát s_0 được lập từ kí hiệu xuất phát S. Các chuyển dịch trong M được tạo từ các sản xuất của G theo cách sau. Chuyển dịch từ s_A đến s_F ứng với đầu vào a sẽ được đưa vào nếu $A \rightarrow a$ là một sản xuất và chuyển dịch từ s_A đến s_B ứng với đầu vào a sẽ được đưa vào nếu $A \rightarrow aB$ là một sản xuất. Tập các trạng thái kết thúc bao gồm s_F và cũng bao gồm cả s_0 nếu $S \rightarrow \lambda$ là một sản xuất trong G. Không mấy khó khăn có thể chứng tỏ được rằng ngôn ngữ được đoán nhận bởi M bằng ngôn ngữ được sinh bởi văn phạm G, tức là L(M) = L(G). Điều này có thể được làm bằng cách xác định các từ dẫn tới một trạng thái kết thúc. Các chi tiết xin dành cho độc giả xem như một bài tập.

Trước khi chứng minh phần đảo lại, ta sẽ minh họa cách xây dựng một máy không tất định đoán nhận cùng một tập như một văn phạm chính quy.

VÍ DỤ 3. Dựng một ô tô mat hữu hạn không tất định đoán nhận ngôn ngữ được sinh bởi văn phạm chính quy G = (V, T, S, P), trong đó V = {0, 1, A, S}, T = {0, 1} và các sản xuất trong T là $S \rightarrow 1A$, $S \rightarrow 0$, $S \rightarrow \lambda$, $A \rightarrow 0A$, $A \rightarrow 1A$ và $A \rightarrow 1$.

Giải: giản đồ trạng thái của ô tô mat không tất định đoán nhận L(G) được cho trên hình 4. Ô tô mat này được xây dựng theo thủ tục được mô tả trong phần chứng minh ở trên. Trong ô tô mat này s_0 là trạng thái tương ứng với S, s_1 là trạng thái tương ứng với A và s_2 là trạng thái kết thúc.



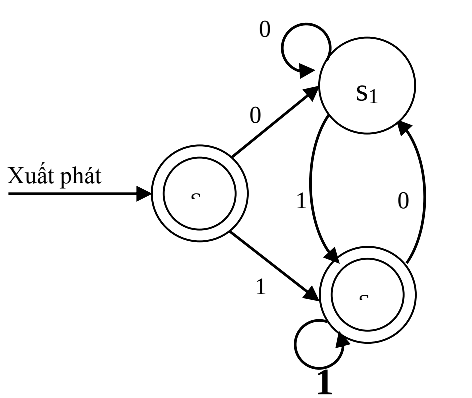
Hình 4. Ôtômat hữu hạn không tất định đoán nhận L(G)

Bây giờ ta sẽ hoàn tất việc chứng minh định lý 2.

Chứng minh. Ta sẽ chứng minh rằng, nếu một tập là chính quy thì tồn tại một văn phạm chính quy sinh ra tập chính quy đó. Giả sử M là một máy hữu hạn trạng thái đoán nhận tập đó với tính chất s_0 – trạng thái xuất phát của M – không bao giờ là trạng thái kế tiếp đối với một chuyển dịch. (Ta sẽ tìm được máy có tính chất này ở bài tập 14). Văn phạm G = (V, T, S, P) được xác định như sau: Tập V các kí hiệu của G được lập bằng cách gán một kí hiệu cho mỗi trạng thái của S và mỗi kí hiệu đầu vào trong I. Kí hiệu xuất phát S là kí hiệu được lập từ trạng thái xuất phát s_0. Tập P các sản xuất trong G được lập từ các chuyển dịch trong M. Đặc biệt, nếu trạng thái s chuyển tới từ một trạng thái kết thúc khi đầu vào là s thì sản xuất $A_s \rightarrow a$ sẽ được bao hàm trong P, ở đây A_s là kí hiệu không kết thúc được tạo từ trạng thái s. Nếu trạng thái s chuyển tới trạng thái t khi đầu vào là a thì sản xuất $A_s \rightarrow aA_t$ sẽ đực bao hàm trong P. Sản xuất $S \rightarrow \lambda$ được bao hàm trong P nếu và chỉ nếu $\lambda \in L(M)$. Vì các sản xuất của G tương ứng với các chuyển dịch trong M và các sản xuất dẫn tới các kí hiệu kết thúc tương ứng với các chuyển dịch đến trạng thái kết thúc, nên không khó khăn gì ta có thể chứng minh được rằng L(G) = L(M). Các chi tiết xin dành cho bạn đọc như một bài tập.

Ví dụ sau minh hoạ cách xây dựng một văn phạm từ một đoạn ôtômat đoán nhận ngôn ngữ sinh bởi văn phạm đó.

VÍ DỤ 4: Tìm một văn phạm chính quy sinh tập chính quy được đoán nhận bởi ôtômat cho trên hình 5.



Hình 5: Một ôtômat hữu hạn

Giải

Văn phạm G = (V, T, S, P) sinh ra tập được đoán nhận bởi ôtômat đó, với V = {S, A, B, 0, 1}, trong đó S, A, B tương ứng với các trạng thái s_0, s_1 và s_2, T = {0, 1}, S là kí hiệu xuất phát và các sản xuất là $S \rightarrow 0A$, $S \rightarrow 1B$, $S \rightarrow 1$, $S \rightarrow \lambda$, $A \rightarrow 0A$, $A \rightarrow 1B$, $A \rightarrow 1$, $B \rightarrow 0A$, $B \rightarrow 1B$ và $B \rightarrow 1$.

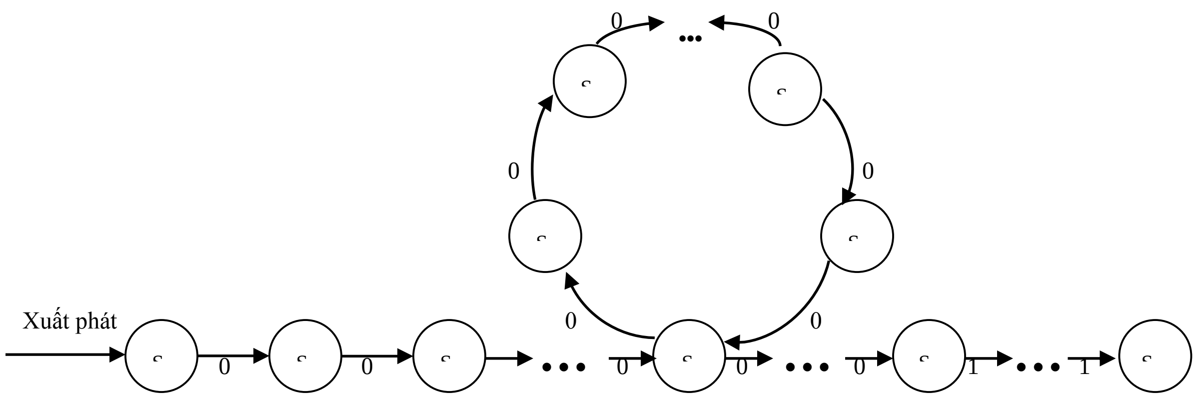
TẬP KHÔNG ĐƯỢC ĐOÁN NHẬN BỞI MỘT ÔTÔMAT HỮU HẠN

Chúng ta đã thấy rằng một tập được đoán nhận bởi một ôtômat hữu hạn nếu và chỉ nếu nó là một tập chính quy. Mặt khác cũng có những tập không phải là chính quy được cho trong các ví dụ sau đây. Kĩ thuật được dung dưới đây mô tả một tập không phải là chính quy là một phương pháp quan trọng để chứng minh một số tập không phải là chính quy.

VÍ DỤ 5: Chứng tỏ rằng tập $\{0^n1^n \mid n = 0, 1, 2,\ldots\}$ tạo bởi tất cả các xâu gồm một khối các số 0 tiếp sau bởi một khối các số 1 với số lượng như nhau là một tập không chính quy.

Giải: Giả sử tập này là chính quy. Khi đó sẽ có một ôtômat hữu hạn tất định $M = (S, I, f, s_0, F)$ đoán nhận tập đó. giả sử N là số trạng thái trong M, tức là $N = | S |$. Vì M đoán nhận tất cả các xâu tạo bởi một số các số 0 và tiếp theo bởi một số như thế các số 1, nên M phải đoán nhận 0^N1^N. Giả sử $s_0, s_1, \ldots, s_{2N}$ là dãy các trạng thái nhận được bắt đầu từ s_0 và dùng các kí hiệu 0^N1^N như một đầu vào, sao cho $s_1 = f(s_0, 0)$, $s_2 = f(s_1, o)$, …, $s_N = f(s_{N-1}, 0)$, $s_{N+1} = f(s_N, 1)$, …, $s_{2N} = f(s_{2N-1}, 1)$. Chú ý rằng s_{2N} là trạng thái kết thúc.

Vì chỉ có N trạng thái, nên theo nguyên lí Dirichlet, ít nhất có hai trạng thái trong số N+1 trạng thái đầu tiên $s_0, s_1, \ldots, s_N$ phải như nhau. Ví dụ s_i và s_j là hai trạng thái đồng nhất đó, với $0 \leq i < j \leq N$. Điều này có nghĩa là $f(s_i, 0^t) = s_j$ với $t = j - i$. Từ đây suy ra rằng vòng dẫn từ s_i quay lại chính nó nhận được bằng cách dùng đầu vào 0 tổng cộng t lần, như được biểu diễn bởi giản đồ trạng thái trên hình 6.



Hình 6: Đường đi tạo bởi 0^N1^N.

Bây giờ ta xét xâu đầu vào $0^N0^t1^N = 0^{N+t}1^N$. Đoạn đầu của xâu này có số các số 0 liên tiếp nhiều hơn số các số 1 tiếp sau t vị trí. Vì xâu này không có dạng 0^n1^n (do có số các số 0 nhiều hơn số các số 1) nên nó không được đoán nhận bởi M. Do đó, $f(s_0, 0^{N+t}1^N)$ không phải là trạng thái kết thúc. Tuy nhiên, khi dùng xâu $0^{N+t}1^N$ như một đầu vào ta sẽ kết thúc ở chính trạng thái s_{2N} như trước. Đó là bởi vì t số 0 thừa trong xâu đó sẽ đi quanh vòng từ s_i trở lại chính nó thêm một lần nữa như trên hình 6. Sau đó phần còn lại của xâu sẽ đưa chúng ta tới đúng trạng thái kết thúc như trước. Mâu thuẫn này chứng tỏ $\{0^n1^n \mid n = 0, 1, 2, \ldots\}$ không phải là tập chính quy.

CÁC LOẠI MÁY MẠNH HƠN

Các ôtômat hữu hạn không thể thực hiện được nhiều tính toán. Hạn chế chủ yếu của các máy này là lượng nhớ hữu hạn của chúng. Chính điều này đã ngăn trở nó đoán nhận các ngôn ngữ không chính quy, như $\{0^n1^n \mid n = 0, 1, 2,\ldots\}$. Tuy nhiên, có một văn phạm phi ngữ cảnh đoán nhận tập này. Một văn phạm như vậy đã được cho trong ví dụ 5 ở mục 11.1.

Vì những hạn chế của các ôtômat hữu hạn, nên cần phải dung các mô hình tính toán khác mạnh hơn. Một mô hình như vậy là ôtômat đẩy xuống (pushdown). Một ôtômat đẩy xuống là một ôtômat hữu hạn kèm theo một băng đẩy xuống tạo một bộ nhớ không hạn chế. Các kí hiệu có thể được đặt vào hoặc lấy ra ở đỉnh của băng đó. Một tập sẽ được một ôtômat đẩy xuống đoán nhận theo hai cách. Thứ nhất, tập sẽ được đoán nhận nếu nó gồm tất cả các xâu tạo ra một băng trống khi các xâu đó được dung làm đầu vào. Thứ hai, tập sẽ được đoán nhận nếu nó gồm tất cả các xâu dẫn tới một trạng thái kết thúc khi các xâu đó được dùng làm đầu vào. Người ta có thể chứng minh được rằng một tập được đoán nhận bởi một ôtômat đẩy xuống nếu và chỉ nếu nó là ngôn ngữ được sinh bởi một văn phạm phi ngữ cảnh.

Tuy nhiên, có những tập không thể biểu diễn được như một ngôn ngữ được sinh bởi một văn phạm phi ngữ cảnh. Một trong những tập như vậy là $\{0^n1^n2^n \mid n=0,1, 2,\ldots\}$. Ta có thể chứng minh tập này không thể được đoán nhận bởi một ôtômát đẩy xuống, nhưng sẽ không chứng minh vì không có đủ các công cụ cần thiết. (Tuy nhiên, một phương pháp chứng minh sẽ được cho trong bài tập 28 của bài tập bổ sung ở cuối chương này). Băng có thể được dùng để cho thấy rằng xâu bắt đầu với các dãy số 0 được tiếp theo bằng các số 1 với số lượng như nhau bằng cách đặt một ký hiệu lên băng đối với số 0 (chừng nào chỉ các số không còn đang được đọc) và lấy đi một trong số các ký hiệu đó đối với mỗi số 1 (chừng nào chỉ những số 1 tiếp theo các số 0 còn đang được đọc). Nhưng một khi điều đó được làm, băng sẽ là trống và không còn cách nào khác để xác định xem trong xâu có số các số 2 đúng bằng số các số 0 hay không

Có những máy khác được gọi là ôtômát tuyến tính giới nội. Các máy này mạnh hơn các ôtômát đẩy xuống, chúng có thể đoán nhận các tập như $\{0^n1^n2^n \mid n=0,1,2,\ldots\}$. Đặc biệt, các ôtômát tuyến tính giới nội có thể đoán nhận cả các ngôn ngữ cảm ngữ cảnh. Tuy nhiên các máy này lại không thể đoán nhận được tất cả các ngôn ngữ được sinh bởi các văn phạm cấu trúc câu. Để tránh những hạn chế của các loại máy trên, người ta sử dụng mô hình được gọi là máy Turing, theo tên nhà toán học Anh Alan Turing. Một máy Turing là một máy ôtômát hữu hạn có kèm theo một băng vô hạn về cả hai phía. Một máy Turing đọc và viết trên băng và nó cỏ thể chuyển động qua lại dọc theo băng đó. Các máy Turing có thể đoán nhận tất cả các ngôn ngữ được sinh bởi các văn phạm cấu trúc câu. Thêm vào đó, các máy Turing có thể mô hình tất cả các tính toán có thể được thực hiện trên một máy tính. Do sức mạnh của mình, các máy Turing đã được nghiên cứu rộng rãi trong tin học lý thuyết. Chúng ta sẽ nghiên cứu một cách vắn tắt các máy này ở mục sau.

Bài tập

1. Mô tả bằng lời các xâu trong các tập chính quy sau:

a) 1*0; b) 1*00*;

c) 111∪001; d) (1∪00)*;

e) (00*1)*; f) (0∪1)(0∪1)*00;

2. Xâu 1011 có thuộc các tập chính quy cho dưới đây không?

a) 10*1*; b) 0*(10∪11)*;

c) 1(01)*1*; d) 1*01(0∪1);

e) (10*(11))*; f) 1(00)*(11)*;

g) (10)*1011; h) (1∪00)(01∪0)1*.

3. Dùng các biểu thức chính quy biểu diễn các tập sau

a) Tập các xâu có một hoặc nhiều hơn số 0 được tiếp sau bởi số 1.

b) Tập các xâu có hai hoặc nhiều hơn ký hiệu được tiếp sau bởi ba hoặc nhiều hơn số 0.

6. Tìm một ôtômát hữu hạn đoán nhận.

a) {λ,0); b) {0,11}; c) {0,11,000}.

7. Dùng các cách dựng được mô tả trong phần chứng minh của định lý Kleene, tìm một ôtômát hữu hạn không tất định đoán nhận các tập sau:

c) Tập các xâu hoặc không có số 1 nào đứng trước một số 0 hoặc không có số 0 nào đứng trước một số 1.

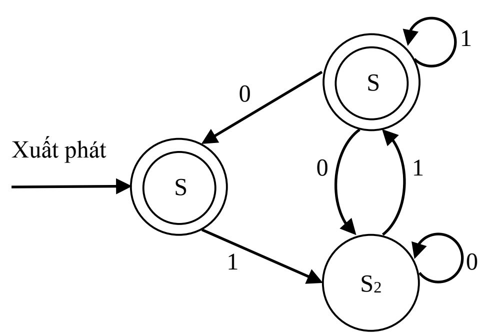
d) Tập các xâu chứa một xâu các số 1 sao cho số các số 1 bằng 2 môđun 3, được tiếp theo bởi một số chẵn các số 0.

4. Dựng một ôtômát hữu hạn tất định đoán nhận các tập sau từ *I**, với *I* là một chữ cái.

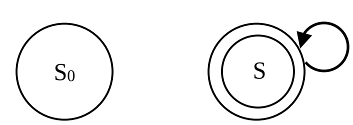
a) ∅; b) {λ}; c) {a} với a∈*I*.

5*. Chứng minh rằng, nếu A là một tập chính quy thì tập tất cả các xâu đảo ngược A^R của các xâu trong A cũng sẽ là chính quy.

10.



11.



a) 0*1*;   b) (0∪11)*;

c) 01*∪00*1.

8. Dựng một ôtômát hữu hạn không tất định đoán nhận các ngôn ngữ được sinh bởi văn phạm chính quy

G= (V,T,S,P), trong đó

V={0,1,S,A,B},

T={0,1}, S là ký hiệu xuất phát và tập P các sản xuất là:

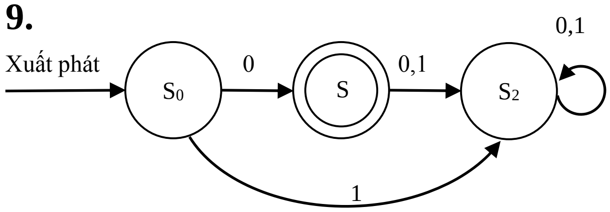
a) $S\rightarrow OA$, $S\rightarrow 1B$, $A\rightarrow 0$, $B\rightarrow 0$;

b) $S\rightarrow 1A$, $S\rightarrow 0$, $S\rightarrow \lambda$, $A\rightarrow 0B$, $B\rightarrow 1B$, $B\rightarrow 1$;

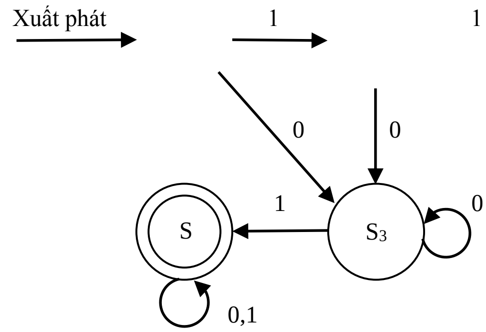
c) $S\rightarrow 1B$, $S\rightarrow 0$, $A\rightarrow 1A$, $A\rightarrow 0B$, $A\rightarrow 1$, $A\rightarrow 0$, $B\rightarrow 1$.

Trong các bài tập 9 – 11, hãy dựng một văn phạm G=(V,T,S,P) sinh ra ngôn ngữ được đoán nhận bởi máy hữu hạn trạng thái đã cho.

9.



16*. Một kỹ thuật quan trọng để chứng minh một tập nào đó là không chính quy được gọi là bổ đề bơm (pumping). Bổ đề bơm phát biểu rằng nếu $M=(S,I,f,s_0,F)$ là một ôtômát hữu hạn



12. Chứng tỏ rằng một ôtômát hữu hạn ấy dựng từ một văn phạm chính quy trong phần chứng minh định lý 2 đoán nhận tập được sinh bởi văn phạm đó.

13. Chứng minh rằng văn phạm chính quy được xây dựng từ một ôtômát hữu hạn trong phần chứng minh của định lý 2 sẽ sinh ra tập được đoán nhận bởi ôtômát đó.

14. Chứng tỏ rằng mỗi một ôtômát hữu hạn tất định đều tương đương với một ôtômát khác có tính chất là trạng thái xuất phát của nó không bao giờ được tái ngộ.

15*. Cho $M=(S, I, f,s_0,F)$ là một ôtômát hữu hạn tất định. Chứng tỏ rằng ngôn ngữ được đoán nhận bởi M, tức L(M), là vô hạn nếu và chỉ nếu có một từ x được đoán nhận bởi M có $l(x)\geq|S|$.

17*. Chứng tỏ rằng tập $\{0^{2n}1^{n} \mid n=0,1,2,,, \}$ là không chính quy. bạn có thể dùng bổ đề bơm cho trong bài tập 16*.

18*. Chứng tỏ rằng tập $\{0^{n^2} \mid n=0,1,2,\ldots \}$ là không chính quy. Bạn có thể dùng bổ đề bơm cho trong bài tập 16.

tất định và nếu x là một xâu thuộc L(M)-ngôn ngữ được đoán nhận bởi M- với $l(x) \geq |S|$, thì có các xâu u, v và w trong *I** sao cho *x=uvw*, $l(uv) \leq |S|$, $l(v) \geq 1$ và $uv^1w \in L(M)$ với i=0,1,2,…Hãy chứng minh bổ đề bơm. (Gợi ý: Dùng chính ýtưởng được dùng trong ví dụ 5).

19*. Chứng tỏ rằng tập các xâu thuận nghịch trên tập {0,1} là không chính quy. Bạn có thể dùng bổ đề bơm cho trong bài tập 16. (Gợi ý: Xét các xâu có dạng 0^N1^N).

20.** Chứng tỏ rằng tập được đoán nhận bởi một ôtômát hữu hạn là một tập chính quy. (Đây là phần *nếu* của định lý Kleene).