形式語言與自動機・期末考

2021 秋　考題回憶版　by BoxWorld

一、判斷題（1@2）

1. 存在可以接受對角語言 L_d 的多棧 PDA。
2. 在一個 DFA 中，設狀態 r 和 s 通過某個輸入符號 a 可分別轉移到 p 和 q（即 $\delta(r,a)=p, \delta(s,a)=q$）則有「$r$ 和 s 可區別 $\Rightarrow$ p 和 q 可區別」。
3. 若 L 是正規語言，a 是字母表中的符號，則 $a\backslash L=\{w|aw \in L\}$ 也是正規語言。
4. 對角語言 L_d 可以歸約到通用語言 L_u 的補語言。
5. 正規表達式 $0+(\varepsilon+1)(\varepsilon+1)^*0$ 可簡化為 1^*0。
6. 圖靈機停機問題是一個 *NP-complete* 問題。
7. 「兩個正規語言是否擁有至少一個公共串」是可判定的。
8. 如果被施加替換（*substitution*）運算的某個語言是上下文無關語言，則進行替換運算之後得到的語言也是上下文無關語言。

二、單項選擇題（1@2）

1. 下列語言中，（）不是遞歸可枚舉語言。

   A. 語言 L_u（課程定義的通用語言）

   B. 語言 L_H（課程中圖靈機停機問題所定義的語言）

   B. 語言 L_H（課程中圖靈機停機問題所定義的語言）的補語言

   D. 以上皆非

2. 下列問題中，（）是可判定的。

   A. 一個遞歸可枚舉語言是否為空

   B. 一個遞歸可枚舉語言是否為正規語言

   B. 一個遞歸可枚舉語言是否為上下文無關語言

   D. 以上皆非

3. 下列語言中，（）不是任何 PDA 的語言。

   A. $\{ww|w \in \{a,b\}^*\}$

   B. $\{ww|w \in \{a,b\}^*\}$ 的補語言

   C. $\{ww^R|w \in \{a,b\}^*, w^R\text{為}w\text{的反向}\}$ 的補語言

   D. 以上皆非

4. 下列語言中，（）是某個 PDPA 的語言。

A. $\{cww^R|w \in \{a,b\}^*, w^R為w的反向\}$

B. $\{wcw^R|w \in \{a,b\}^*, w^R為w的反向\}$

C. $\{ww^Rc|w \in \{a,b\}^*, w^R為w的反向\}$

D. 以上皆非

5. 下列語言中，（）是某個空棧接受的 PDPA 的語言。

A. $\{ww^R|w \in \{a,b\}^*, w^R為w的反向\}$ 的補語言

B. $\{wcw^R|w \in \{a,b\}^*, w^R為w的反向\}$

C. $\{wcw^R|w \in \{a,b\}^*, w^R為w的反向\}$ 的補語言

D. 以上皆非

6. 下列語言中，（）不是任何 DFA 的語言

A. $\{wxw^R|w,x \in \{a,b\}^*, w^R為w的反向\}$

B. $\{wxw^R|w,x \in \{a,b\}^*, w^R為w的反向\}$ 的補語言

C. $\{wxw|w,x \in \{a,b\}^*\}$

D. 以上皆非

三、簡答題

1. 【5 分】設 CFG $G = (\{S,A,B,C,D\},\{a,b,c\},P,S)$，其中 P 由下列產生式構成：

$$
\begin{aligned}
S &\to AB|BC|\varepsilon \\
A &\to AB|C \\
B &\to Bb|a \\
C &\to \varepsilon|Cc \\
D &\to B|\varepsilon
\end{aligned}
$$

(a) 消去 P 中的 ε - 產生式得到產生式集合 P_1，構成 CFG G'，使得 $L(G') = L(G) - \{\varepsilon\}$，求 P_1。

(b) 消去 P_1 中的 Unit 產生式得到產生式集合 P_2，構成 CFG G''，使得 $L(G'') = L(G')$，求 P_2。

(c) 消去 P_2 中的無用符號得到產生式集合 P_3，構成 CFG G'''，使得 $L(G''') = L(G'')$，求 P_3。

(d) 根據 P_3 的結果，添加合適的非終結符，構造滿足 Chomsky 範式要求的產生式集合 P_4，使得結果文法的語言與 $L(G''')$ 相等，求 P_4。

2. 【4 分】文法 G（S 為開始符號）的產生式集合為：

$$
\begin{aligned}
S &\to AB \\
A &\to SS|SA|b \\
B &\to BA|b|a
\end{aligned}
$$

X_{13}		
X_{12}	X_{23}	
X_{11}	X_{22}	X_{33}
b	a	b

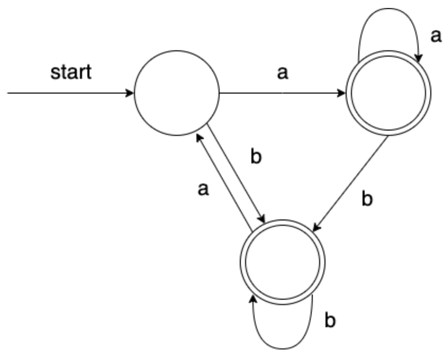
上圖表示對於文法 G 和字符串 bab 應用 CYK 算法時所構造的表。

(a) 分別計算圖中所有 X_{ij} $(1 \le i, j \le 3)$。

(b) 是否有 $bab \in L(G)$?

3. 【6 分】下圖表示一個有限自動機 A:

（注：本題中多處出現有限自動機的描述，可以是也可以不是 DFA）



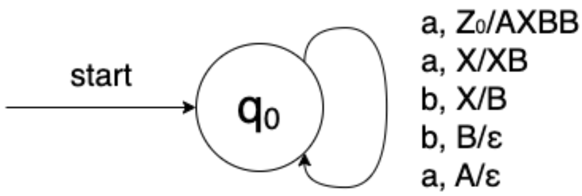
(a) 試採用課程中所介紹的方法，給出一個有限自動機 B，使得 $L(B) = (L(A))^R$。（$(L(A))^R$ 為 $L(A)$ 的反向）

(b) 試採用課程中的方法，給出一個有限自動機 C，使得 $L(C) = \{a, b\}^* - L(A)$。

(c) 設映射 $h : \{0, 1\} \to \{a, b\}^*$ 定義為 $h(0) = ab, h(1) = ba$;

試構造一個有限自動機 D，使得 $L(D) = h^{-1}(L(A))$。

4. 【6 分】考慮如下 PDA 狀態轉移圖:



該圖刻畫了 PDA $P = (\{q_0\}, \{a, b\}, \{Z_0, X, A, B\}, \delta, q_0, Z_0)$ 的轉移規則。

(a) 試嚴格利用課程中介紹的從空棧接受的 PDA 到 CFG 的轉換算法，定義一個與該 PDA 等價的 CFG，開始符號設為 S。

(b) 對於該 PDA 接收的語言 $L(P)$，是否有

$$L(P) \subseteq \{w | w = a^n b^m, \text{其中 } 0 \le m \le 2n \le 4m\}$$

(c) 該 PDA 接收的語言 $L(P)$ 是否為正規語言？結出結論即可。

5. 【6 分】若 G 為包含 p 個產生式的上下文無關文法，其中每個產生式的長度小於等於 n。假設存在推導 $A \Rightarrow_G^* \varepsilon$ 則對於該推導存在推導步數的上界，使得一定存在一種推導方法的推導步數小於等於該上界。給定 n 和 p （$n, p \geq 1$），求出這一上界，並對於這一上界，給出推導步數達上界的上下文無關文法。

例： 若 $n = 2, p = 2$ ，對應上界為 3，一個對應的 CFG 為 $A \Rightarrow BB, B \Rightarrow \epsilon$。

注：不需要對推導步數的上界給出對應的證明。

6. 【4 分】定義兩個語言 L 和 M 的對稱差 $SD(L, M)$ 為所有僅被語言 L 或 M 中的一個所包含的字符串組成的集合。例如，若 $L = \{aa, bb\}, M = \{bb, aab\}$，則 $SD(L, M) = \{aa, aab\}$。

   (a) 給出 $SD(L, M)$ 的形式化描述。例：L^+ 的形式化描述為 $L^+ = L^1 \cup L^2 \cup L^3 \cup \ldots$

   (b) 若 L, M 均為 CFL，$SD(L, M)$ 是否一定是 CFL？給出證明或給出反例。

7. 【4 分】對於語言

$$L = \{ss^Rs | s \in \{0,1\}^*, s^R \text{為} s \text{的反向}\}$$

可以利用 Pumping 引理證明 L 不是上下文無關語言，以下是一個證明概要：

對於任意的 $n \geq 1$，取 $s = ① \in \{0,1\}^*$，令 $z = ss^Rs$，則 $z \in L$。
對任意滿足條件 $z = uvwxy \wedge vx \neq \varepsilon \wedge |vwx| \leq n$ 的 u, v, w, x, y，取 $k = ②$，有 $uv^kwx^ky \notin L$。

試在其中 ① 和 ② 處填寫適當的內容。

四、設計題

1. 【5 分】構造接受下列語言 L 的一個有限自動機（DFA，NFA，ε-NFA 均可），要求狀態數不超過 8， 且用狀態轉移圖的方式給出答案：

$$L = \{w | w = a^n b^m,\ m, n \geq 0\ , \text{且} \ w \ \text{中既不包含子串} \ aaabb\text{，也不包含子串} \ aabbbb\}$$

注：要求狀態數不超過 8，並不意味狀態數一定會達到 8，後面的題目亦然。

2. 【5 分】給出下列正規語言 L 的一個正規表達式：

$$L = \{w | w = a^n b^m,\ m, n \geq 0\ , \text{且} \ w \ \text{中既不包含子串} \ aaabb\text{，也不包含子串} \ aabbbb\}$$

3. 【5 分】給出下列語言 L 的一個上下文無關文法 G，要求 G 的非終結符只有一個 S。用你的文法 G 驗證：對於任意串 $w = a^n b^m \in L$，w 可被 G 接受。

$$L = \{w | w = a^n b^m, \text{其中} \ 0 \leq m \leq 2n \leq 4m\}$$

4. 【5 分】試構造接受下列語言 L 的一個 PDA（終態接受和空棧接受均可，必要時給出設計思路），要求該 PDA 的堆棧符號數不超過 3，且用狀態轉移圖描述你的設計：

$$L = \{w | w \in \{a, b\}^*, \text{其中} \ |w| \ \text{為偶數，且} \ w \ \text{不為} \ a^n b^n \ \text{的形式，其中} \ n \geq 0\}$$

5. 【5 分】試設計一個可停機圖靈機 $M = (Q, \{0,1\}, \{0,1..,B\}, \delta, q_0, B, \{q_f\})$ 可以將串 $w \in \{0,1\}^*$ 作為輸入，當到達終態 q_f 時，帶上的內容為將 w 從小到大排序後的結果。**例：** 如輸入串為 101010，則到達終態時，帶上的內容應為 000111。如輸入串為 1000，則到達終態時，帶上的內容應為 0001。如輸入串全為 0 或全為 1，則到達終態時帶上的內容不變。該圖靈機的狀態數不超過 7。到達 q_f 時，對讀寫頭在何處不作要求。用狀態轉移圖描述你所設計的圖靈機。

五、證明題

- 要求證明嚴謹，步驟明確。

1. 【4 分】已知語言 $L_{01} = \{0^k 1^k | k \geq 0\}$ 不是正規語言，試利用該結論及正規語言的封閉運算，證明如下語言 L 不是正規語言：

$$L = \{w | w = a^n b^m, \text{其中 } 0 \leq 3n = 2m\}$$

2. 【6 分】設有語言 $L = \{w | w = a^n b^m, \text{其中 } 0 \leq m \leq 2n \leq 4m\}$。試問，$L$ 是否存在一個非有窮子集是正規語言？如存在這樣的非有窮子集，請給出一個例子；如不存在，請給出證明。

3. 【5 分】證明若 L 為上下文無關語言，R 為正規語言，則 L 和 R 的交 $L \cap R$ 為上下文無關語言。

   提示： 考慮 L 對應的 PDA $P = (Q_P, \Sigma, \Gamma, \delta_P, q_p, Z_0, F_P)$ ，和 R 對應的 DFA $A = (Q_A, \Sigma, \delta_A, q_A, F_A)$，構造 $L \cap R$ 對應的自動機並進行證明。

六、附加題

- 5 分，直接加入總評成績；共兩題，任選一題。
- **注意：** 附加題只有能呈現出核心思路才有可能得到部分分數，建議大家在前面題目已做完且進行充分檢查之後，再看是否有時間考慮下列題目之一。

1. 對於 CFG $G = (V, T, P, S)$，我們先定義如下兩個概念：活前綴（*viable prefix*）和增廣文法（*augmented grammar*）。

   若 $S \Rightarrow^*_{rm} \alpha A w$ 且 $A \Rightarrow \beta$，其中 $\alpha, \beta \in (V \cup T)^*$，$w \in T^*$，$\Rightarrow^*_{rm}$ 表示最右推導（多步），則 $\alpha\beta$ 的任何前綴 γ 都是文法 G 的活前綴。

   若增加產生式 $S' \to S$，其中 $S' \notin V \cup T$，得到 G 的增廣文法 $G = (V, T, P, S')$。

   現在，針對增廣文法 G'，集合 *Prefix* 可歸納定義如下：

   (a) 令 $S \in Prefix$；
   (b) 若 $v \in Prefix$，則 v 的任一前綴 u 都滿足 $u \in Prefix$；
   (c) 若 $v \in Prefix$，且 v 中至少包含一個非終結符，即可以將 v 寫成 $\alpha B \gamma$，其中 B 為非終符。若有產生式 $B \to \beta$ ，則 $\alpha\beta$ 的任一前綴 u 都滿足 $u \in Prefix$；
   (d) *Prefix* 中的元素只能通過上述步驟產生。

   試證明，*Prefix* 可以表示增廣文法 G' 所有活前綴的集合。

2. 在空棧接受的擴展 PDA 基礎上，我們引入一種便於自底向上分析的自動機。一個歸約自動機（*Recursive Automaton*，簡稱 *RA*），是一個六元組：

$$R = (Q, \Sigma, \Gamma, \delta, q_0, Z_0)$$

其中，Q, Σ, Γ 及 q_0 與 PDA 的含義一致，但 δ 定義為：

$$\delta = Q \times (\Sigma \cup \{\varepsilon\}) \times (\Gamma - \{Z_\varepsilon\})^* \to 2^{Q \times \Gamma}$$

$Z_\varepsilon \in \Gamma$ 是終棧符（*end stack symbol*）。定義與 PDA 中類似的 ID，形如（q, w, γ）。不同的是，棧頂在最右邊。定義 ID 間的二元關係 $\vdash_R$：

$$(q, aw, \gamma\alpha) \vdash_R (p, w, \gamma X) \text{ 當且僅當 } (p, X) \in \delta(q, a, \alpha)$$

其中，$p, q \in Q$，$a \in \Sigma \cup \{\varepsilon\}$，$w \in \Sigma^*$，$X \in \Gamma$，$\alpha, \gamma \in \Gamma^*$。

上述 *ID* 間二元關係 $\vdash_R$ 的自反傳遞閉包表示為 $\vdash_R^*$。定義 R 的語言

$$L(R) = \{w | (q_0, w, \varepsilon) \vdash_R^* (q, \varepsilon, Z_\varepsilon), \text{ 其中}, q \in Q\}$$

試證明：對任一歸約自動機 $R = (Q, \Sigma, \Gamma, \delta, q_0, Z_\varepsilon)$，存在某個上下文無關文法 G，使得 $L(G) = L(R)$。