形式语言与自动机 大作业 2

综合练习

1951112　林日中

第 1 题

设

$$\Sigma = \left\{ \begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1\\0\end{bmatrix}, \begin{bmatrix}1\\1\end{bmatrix} \right\}$$

Σ 包括所有高度为 2 的 0 和 1 的列，Σ 上的字符串给出两行 0 和 1. 把每一行看作一个二进制数，令

$$E = \left\{ \omega \in \Sigma^* \middle| \omega \text{ 的下一行是上一行的反转} \right\}$$

例如

$$\begin{bmatrix}0\\1\end{bmatrix} \begin{bmatrix}1\\0\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} \begin{bmatrix}1\\0\end{bmatrix} \in E, \text{while} \begin{bmatrix}0\\0\end{bmatrix} \begin{bmatrix}1\\1\end{bmatrix} \begin{bmatrix}1\\1\end{bmatrix} \begin{bmatrix}0\\1\end{bmatrix} \notin E$$

证明 E 不是正则的.

证明. 假设这个语言 E 是正则的. 那么由正则语言的泵引理，存在一常数 $p > 0$，对于语言 E 中每个满足 $|w| \geq p$ 的字符串 w，存在一组 x, y, z 使得 $w = xyz$ 且有 $|xy| \leq p, |z| \geq 1$，并且有 $\forall k \geq 0, xy^k z \in E$.

考察串 $w = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^p \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]^p, |w| = 2p \geq p$，其中 $1^p 0^p = (0^p 1^p)^R$. 按照泵引理，对串 w 生成一组分解 x, y, z. 由于 $|xy| \leq p$，那么必有泵 $y \in \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^+$.

那么，由语言 E 是正则的，有 $w' = xy^2 z = \left[\begin{smallmatrix}0\\1\end{smallmatrix}\right]^{p+1} \left[\begin{smallmatrix}1\\0\end{smallmatrix}\right]^p$，显然有 $0^{p+1} 1^p \neq (1^{p+1} 0^p)^R$，即 $w' \notin E$.

因此，语言 E 中存在不满足正则语言的泵引理的串，这与它是正则语言的假设矛盾. 故这个语言 E 不是正则的.

第 2 题

证明语言 $F = \{a^i b^j | \exists k \in \mathbb{Z}_+, i = kj\}$ 不是上下文无关的.

证明. 假设语言 F 是上下文无关的. 那么由上下文无关语言的泵引理，存在一常数 $p_0 > 0$，对于语言 F 中每个满足 $|z| \geq p_0$ 的字符串 z，存在一组 u, v, w, x, y 使得 $z = uvwxy$ 且有 $|vwx| \leq p_0, vx \neq \varepsilon$ 且 $\forall i \geq 0, uv^iwx^iy \in F$.

考察串 $a^{p^2}b^p, p > p_0$，对应地有 $k = p$. 对于串 vwx，有 $|vwx| \leq p_0 < p$ 且 $vwx \in a^+, b^+$ or a^+b^+.

当 $vwx \in a^+$ 时,对于串 $uv^0wx^0y = a^{p^2-|vx|}b^p, 0 < |vx| \leq |vwx| < p$,有 $p^2 - p < p^2 - |vx| < p^2$. 显然，因为区间 $(p-1, p)$ 内不可能有整数，故 $p^2 - |vx|$ 不可能被 p 整除. 故此时，有 $uv^0wx^0y \notin F$.

当 $vwx \in b^+$ 时，有 $uv^iwx^iy = a^{p^2}b^{p+i\times|vx|}$. 因为 $\forall i > p^2, p + i \times |vx| > p^2$，故有 $uv^iwx^iy \notin F, i > p^2$.

当 $v \in a^+b^+$ 或 $x \in a^+b^+$ 时，uv^iwx^iy 中 a 与 b 的顺序被打乱，此时有 $uv^iwx^iy \notin F$.

当 $v \in a^+$ 且 $x \in b^+$ 时，对于串 $uv^iwx^iy = a^{p^2+i\times|v|}b^{p+i\times|x|}, |v|, |x| > 0$，有

$$f(i) = \frac{|a|}{|b|} = \frac{p^2 + i \times |v|}{p + i \times |x|}$$

对于 $f(i)$ 的一阶导数 $f'(i)$，因为 $|vwx| \leq p$，有

$$f'(i) = \frac{p \times |v| - p^2}{(p + |x| \times i)^2} < 0$$

故函数 $f(i) \in \left(\frac{|v|}{|x|}, f(0)\right]$，且为单调递减函数，其中

$$\lim_{i\to+\infty} \frac{p^2 + i \times |v|}{p + i \times |x|} = \frac{|v|}{|x|}$$

然而，在区间 $\left(\frac{|v|}{|x|}, p\right]$ 内，只有少于 $f(0) = p$ 个整数，且 $f(i)$ 单调递减；即只对有限的 i 的取值而言，$f(i)$ 是整数. 那么，我们总能找到使 $f(i)$ 不为整数的 i 的取值，此时有 $uv^iwx^iy \notin F$.

综上所述，语言 $F = \{a^ib^j | \exists k \in \mathbb{Z}_+, i = kj\}$ 不是上下文无关的.

第 3 题

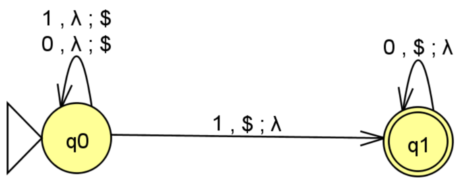
设 $\Sigma = \{0,1\}$，B 为后一半中至少包含一个 1 的串的集合. 换句话说，$B = \{uv | u \in \Sigma^*, v \in \Sigma^*1\Sigma^*, |u| \geq |v|\}$.

a. 设计一个能识别语言 B 的 PDA；

b. 设计一个能产生语言 B 的 CFG.

解答.

a. 能识别语言 B 的 PDA 如下图所示.



b. 构建能产生语言 B 的 CFG $C=(V,\Sigma,R,S)$，其中

$$V=\{A,B\},\Sigma=\{0,1\},S=A$$

R 是由以下产生式组成的集合：

$$A\to 0A0\mid 0B1\mid 1A0\mid 1B1$$

$$B\to 0B0\mid 0B1\mid 1B0\mid 1B1\mid 0\mid 1\mid \varepsilon$$

在这个语法中，我们构建字符串的方式是总在两边各加一个 0 或一个 1，所以我们在跟踪我们字符串的中间点. 如果我们在推导的某个地方得到了 B 变量，我们就可以终止了. 如果我们在字符串的后半部分加上一个 1，我们就能在推导的某个地方得到 B 变量. 然后我们可以继续构建我们的字符串，最终选择 0、1 或 ε 作为我们的中间符号.

第 4 题

证明语言 $L\subseteq\{0\}^*$ 是上下文无关的当且仅当 L 是正则的.

证明. **先证充分性.** 要证充分性，只需证所有正则语言都是上下文无关的. 由乔姆斯基层次结构（Chomsky's Hierarchy），每个高型的语言都是低型的，故每一个正则语言都是上下文相关的. 那么，若 L 是正则的，那么 L 是上下文相关的. 充分性得证.

再证必要性. 已知语言 $L\subseteq 0^*$ 是上下文无关的. 由上下文无关语言的泵引理，存在泵长度 p. 令 $L=L_1\cup L_2$，其中 L_1 是由 L 中长度小于 p 的串的集合，L_2 是由 L 中长度大于等于 p 的串的集合. 同样地，由泵引理，易得 L_1 是有限集，故 L_1 是正则语言，从而只需证 L_2 也是正则语言.

参照泵引理，任取 $z=uvwxy=0^{|uwy|+|vx|}\in L_2,|vx|\le p$. 那么有

$$uv^kwx^ky=0^{|uwy|+k\cdot|vx|}=0^{r+\hat{k}q},q=|vx|,0\le r<q,\hat{k}=\left\lfloor k+\frac{|uwy|}{q}\right\rfloor\ge\left\lfloor\frac{|uwy|}{q}\right\rfloor$$

接下来，我们构造一组 p 个集合 $M_q=\{0^{r+kq}|0\le r<q,k\in\mathbb{N}^*\},1\le q\le p$，并取 $L_{2q}=L\cap M_q$，也即去除 k 取较小值时可能不在 L 中的元素. 我们可以断言

$$L_2=\bigcup_{q=1}^{p}L_{2q}$$

原因有二：

第一，由 L_{2q} 的定义有 $L_2=\bigcup_{q=1}^{p}L_{2q}\subseteq L_2$；

第二，假设 $\exists z\in L_2\cap\bigcup_{q=1}^{p}L_{2q}\subseteq L_2$，参照泵引理做划分，有 $z=uvwxy$，那么有

$$uv^kwx^ky=0^{(|uwy|\bmod |vx|)+\left(k+\left\lfloor\frac{|uwy|}{q}\right\rfloor\right)\cdot|vx|}\in L_2,|vx|\notin\{1,2,\cdots,p\}$$

这与 $|vx|\le p$ 矛盾.

由于 L_{2q} 与正则表达式 $(0|1|2|\cdots|q-1)+q^*$ 表示的语言相差有限个串（因为 0^{r+kq} 中的余数 r 只有有限个取值，即上述正则表达式只有有限次并运算），那么 L 是有限个正则语言的并，故 L 为正则语言.

综上所述，语言 $L\subseteq\{0\}^*$ 是上下文无关的当且仅当 L 是正则的.

第 5 题

二维图灵机具有通常的有穷控制，带却是在所有方向上都是无穷的二维单元网格. 和往常一样，输入放在网格的一行上，带头在输入的左端，控制处于初始状态. 同样和往常一样，以进入终结状态方式来接受. 证明：二维图灵机与普通图灵机接受的语言是一样的.

证明. 为了证明具有双重无限磁带的图灵机（即题中提到的二维图灵机） D 等同于普通图灵机 M 之间的等价性，有必要说明以下两点：

- 任何能被 M 识别的语言 L，也能被 D 识别.
- 任何能被 D 识别的语言 L，也能被 M 识别.

这可以通过“模拟”来说明，即模拟 D 像 M 一样工作，反之亦然.

(i) 用 D 模拟 M.

a) 在 D 的左手边做标记.

b) 防止 D 把它的头移到标记的左边.

这样一来，D 就模拟了 M.

(ii) 用 M 模拟 D.

为了用一个普通图灵机 M 来模拟具有双重无限磁带的图灵机 D,可以用一个 2 个带子的图灵机来模拟它（由 Hopcroft 课本 8.4.2 小节的定理 8.9，有 2 个带子的图灵机与普通图灵机等价）.

因此，我们用 2 个带子的图灵机 M_1 模拟具有双重无限磁带的图灵机 D.

M_1 的第一条带子上写有输入字符串，第二条带子是空白的.

现在,在输入字符串的起始单元,将具有双重无限磁带的图灵机 D 的带子切成两部分.

- 第一部分包含输入字符串和其右边的所有空白.
- 第二部分包含输入字符串的左边出现在第二个磁带上，顺序相反.

这样一来，M 就模拟了 D.

综合 (i) 和 (ii),有 D 等价于 M,即具有双重无限磁带的图灵机等同于普通图灵机. 也即二维图灵机与普通图灵机接受的语言是一样的.

第 6 题

课程教学质量评估：

a. 请对你个人的学习效果进行评价（分成优、良、中、及格、不及格五个档次），简要说明理由；

b. 请对本课程的教学方式、教学内容、教学安排等教学环节进行简要评价，提出宝贵的意见和建议；

c. 个人作品展示，包括与本课程直接相关的思考探索、定理证明、软件作品等. 如涉及较多文件，请压缩后提交.

解答.

a.

优.

我对课程知识理解程度比较好. 我在修读《形式语言与自动机》课程前，已经修读了它的后续课程《编译原理》，因此我在开课前对形式语言与自动机理论已有一定的了解，但仅局限于在《编译原理》课程中使用到的那部分. 在本课程期间，我系统地学习了形式语言与自动机理论中的有穷自动机、正则表达式、正则语言、上下文无关语言、下推自动机、图灵机等知识和与它们相关的理论知识，不仅回顾了编译过程中的初始化方法、词法分析方法运

用到的自动机相关知识,还进一步从理论的角度重新深刻地学习了形式语言与自动机知识系统.

我的课堂参与度比较高. 我从未迟到或缺勤，即使囿于在线课程的局限性，也坚持认真听课，紧跟老师的授课步伐，没有丝毫懈怠. 在课上和课下，我积极以课堂交流、课程群交流、私信交流等形式与老师、同学们探讨理论知识，按时、按质、按量提交课程作业，勇猛精进，踔厉奋发.

b.

由于疫情的影响，课程的绝大部分只能在线上进行，教学形式被限制，缺乏师生之间的眼神交流、面对面即时问答等直接的互动，但课程组织得很好，课程内容丰富，课程评价体系完善，我对课程的感觉很好，在课程期间收获很多.

我的建议是：老师可以再多介绍一些计算理论相关的知识，还可以稍微降低课后作业的强度和难度.

c.

我使用 Qt C++ 语言实现了一个类 C 语言编译器，可以提供词法分析、语法分析、符号表管理、中间代码生成以及目标代码生成等功能，已打包为 `1951112-林日中-类C编译器.zip`.