计算理论导论

习题九: 可归约性 (2)

中国人民大学　信息学院　崔冠宇　2018202147

1. **5.21** Let $AMBIG_{\mathsf{CFG}} = \{\langle G\rangle \mid G$ is an ambiguous CFG$\}$. Show that $AMBIG_{\mathsf{CFG}}$ is undecidable. (Hint: Use a reduction from *PCP*. Given an instance

$$P = \left\{ \left[\frac{t_1}{b_1}\right], \left[\frac{t_2}{b_2}\right], \cdots, \left[\frac{t_k}{b_k}\right] \right\}$$

of the Post Correspondence Problem, construct a CFG G with the rules

$$S \to T \mid B$$

$$T \to t_1 T \mathtt{a}_1 \mid \cdots \mid t_k T \mathtt{a}_k \mid t_1 \mathtt{a}_1 \mid \cdots \mid t_k \mathtt{a}_k$$

$$B \to b_1 B \mathtt{a}_1 \mid \cdots \mid b_k B \mathtt{a}_k \mid b_1 \mathtt{a}_1 \mid \cdots \mid b_k \mathtt{a}_k,$$

where $\mathtt{a}_1, \cdots, \mathtt{a}_k$ are new terminal symbols. Prove that this reduction works.)

证明: 归约函数按照上面定义，下面只证明归约的正确性。

1. $P \in \textit{PCP} \Rightarrow f(P) \in AMBIG_{\mathsf{CFG}}$:

   因为 $P \in \textit{PCP}$，根据 *PCP* 的定义，存在 $i_1, i_2, \cdots, i_k$ 满足 $t_{i_1} t_{i_2} \cdots t_{i_k} = b_{i_1} b_{i_2} \cdots b_{i_k}$。考虑下面两个推导：

   - $S \Rightarrow T \Rightarrow t_{i_1} T \mathtt{a}_{i_1} \Rightarrow t_{i_1} t_{i_2} T \mathtt{a}_{i_2} \mathtt{a}_{i_1} \Rightarrow \cdots \Rightarrow t_{i_1} t_{i_2} \cdots t_{i_{k-1}} T \mathtt{a}_{i_{k-1}} \cdots \mathtt{a}_{i_2} \mathtt{a}_{i_1}$

     $\Rightarrow t_{i_1} t_{i_2} \cdots t_{i_{k-1}} t_{i_k} \mathtt{a}_{i_k} \mathtt{a}_{i_{k-1}} \cdots \mathtt{a}_{i_2} \mathtt{a}_{i_1}$

   - $S \Rightarrow B \Rightarrow b_{i_1} T \mathtt{a}_{i_1} \Rightarrow b_{i_1} b_{i_2} T \mathtt{a}_{i_2} \mathtt{a}_{i_1} \Rightarrow \cdots \Rightarrow b_{i_1} b_{i_2} \cdots b_{i_{k-1}} T \mathtt{a}_{i_{k-1}} \cdots \mathtt{a}_{i_2} \mathtt{a}_{i_1}$

     $\Rightarrow b_{i_1} b_{i_2} \cdots b_{i_{k-1}} b_{i_k} \mathtt{a}_{i_k} \mathtt{a}_{i_{k-1}} \cdots \mathtt{a}_{i_2} \mathtt{a}_{i_1}$

   由于 $t_{i_1} t_{i_2} \cdots t_{i_k} = b_{i_1} b_{i_2} \cdots b_{i_k}$，可见该字符串有两个不同的最左推导，于是 $f(P)$ 有歧义。

2. $P \notin \textit{PCP} \Rightarrow f(P) \notin \textit{AMBIG}_{\mathsf{CFG}}$，转化为证明 $f(P) \in \textit{AMBIG}_{\mathsf{CFG}} \Rightarrow P \in \textit{PCP}$：

   容易看出，一个字符串在 $G[T]$ 中最多存在一种推导，在 $G[B]$ 中也是，于是若 $f(P)$ 有歧义，则两个不同的推导一定是某一个第一次推导使用了 $S \to T$，另一个使用了 $S \to B$。设某个有两种不同推导的字符串 w 有后缀 $\mathsf{a}_{i_k}\mathsf{a}_{i_{k-1}}\cdots\mathsf{a}_{i_1}$，则 $i_1, i_2, \cdots, i_k$ 是 *PCP* 的一个解。

综上，归约是正确的。

2. **5.30** Use Rice's theorem, which appears in Problem 5.28, to prove the undecidability of each of the following languages.

a. $\textit{INFINITE}_{\mathsf{TM}} = \{\langle M \rangle \mid M \text{ is a TM and } L(M) \text{ is an infinite language}\}$.

b. $\{\langle M \rangle \mid M \text{ is a TM and } \texttt{1011} \in L(M)\}$.

c. $\textit{ALL}_{\mathsf{TM}} = \{\langle M \rangle \mid M \text{ is a TM and } L(M) = \Sigma^*\}$.

证明:

a. 分别验证 Rice's Theorem 的两个条件：

1. $\textit{INFINITE}_{\mathsf{TM}}$ 是一个性质 (property)：对于两个 TM M_1 和 M_2，若 $L(M_1) = L(M_2)$，则 $\langle M_1 \rangle \in \textit{INFINITE}_{\mathsf{TM}} \Leftrightarrow L(M_1)$ 为无限集 $\Leftrightarrow L(M_2)$ 为无限集 $\Leftrightarrow \langle M_2 \rangle \in \textit{INFINITE}_{\mathsf{TM}}$。

2. $\textit{INFINITE}_{\mathsf{TM}}$ 是非平凡 (non-trivial) 的：容易找到 TM M_1，满足 $L(M_1) = \emptyset$（定义见下），以及 TM M_2，满足 $L(M_2) = \Sigma^*$（定义见下）。这样 $\langle M_1 \rangle \notin \textit{INFINITE}_{\mathsf{TM}}$，且 $\langle M_2 \rangle \in \textit{INFINITE}_{\mathsf{TM}}$。

几个图灵机的定义：

$M_1 =$"对于任意输入 w，拒绝。"

$M_2 =$"对于任意输入 w，接受。"

由 Rice's Theorem，结论得证。

b. 记 $L = \{\langle M \rangle \mid M \text{ is a TM and } 1011 \in L(M)\}$。分别验证 Rice's Theorem 的两个条件：

1. L 是一个性质 (property)：对于两个 TM M_1 和 M_2，若 $L(M_1) = L(M_2)$，则 $\langle M_1 \rangle \in L \Leftrightarrow 1011 \in L(M_1) \Leftrightarrow 1011 \in L(M_2) \Leftrightarrow \langle M_2 \rangle \in L$。

2. L 是非平凡 (non-trivial) 的：容易找到 TM M_1，满足 $1011 \notin L(M_1)$（定义见下），以及 TM M_2，满足 $1011 \in L(M_2)$（定义见下）。这样 $\langle M_1 \rangle \notin L$，且 $\langle M_2 \rangle \in L$。

几个图灵机的定义：

$M_1 =$“对于任意输入 w，拒绝。”

$M_2 =$“对于任意输入 w，如果 $w = 1011$，接受；否则拒绝。”

由 Rice's Theorem，结论得证。

c. 分别验证 Rice's Theorem 的两个条件：

1. ALL_{TM} 是一个性质 (property)：对于两个 TM M_1 和 M_2，若 $L(M_1) = L(M_2)$，则 $\langle M_1 \rangle \in L \Leftrightarrow L(M_1) = \Sigma^* \Leftrightarrow L(M_2) = \Sigma^* \Leftrightarrow \langle M_2 \rangle \in L$。

2. L 是非平凡 (non-trivial) 的：容易找到 TM M_1，满足 $L(M_1) = \emptyset$，以及 TM M_2，满足 $L(M_2) = \Sigma^*$。这样 $\langle M_1 \rangle \notin L$，且 $\langle M_2 \rangle \in L$。

由 Rice's Theorem，结论得证。

3. Which of the following are decidable and which are undecidable? Give proofs. Use Rice's Theorem whenever possible in showing undecidability.

a) $\{\langle M \rangle \mid M$ is a Turing machine and M has more than 30 states. $\}$.

b) $\{\langle M \rangle \mid M$ is a Turing machine and $L(M)$ is recognized by some Turing machine M' with more than 30 states. $\}$.

c) $\{\langle M\rangle \mid M$ is a Turing machine that accepts every word of the form 0^n1^n, $n \geq 0$ (and possibly other words). $\}$.

d) $\{\langle M, N\rangle \mid M$ and N are Turing machines and $L(M) \subseteq L(N) \cup A$, where $A = 0^*$, the set of finite strings consisting of just 0s$\}$.

证明:

a) 这个语言（记作 L_1）是可判定的，构造一台 TM D 判定 L_1：

$D =$“对于输入 $w = \langle M\rangle$： 检查 $\langle M\rangle$ 中状态数，若大于 30，接受；否则拒绝。”

b) 这个语言（记作 L_2）是可判定的，因为所有 $\langle M\rangle$ 都满足这个性质：

1. 若 M 本身多于 30 个状态（可以利用 **a)** 判定），则取 $M' = M$ 即可；
2. 否则在 M 中增加若干不在原状态中的无用状态，使新图灵机 M' 多于 30 个状态即可。

c) 这个语言（记作 L_3）是不可判定的，用 Rice's Theorem。分别验证 Rice's Theorem 的两个条件：

1. L_3 是一个性质 (property)：对于两个 TM M_1 和 M_2，若 $L(M_1) = L(M_2)$，则 $\langle M_1\rangle \in L_3 \Leftrightarrow \{0^n1^n | n \geq 0\} \subseteq L(M_1) \Leftrightarrow \{0^n1^n | n \geq 0\} \subseteq L(M_2) \Leftrightarrow \langle M_2\rangle \in L$。
2. L_3 是非平凡 (non-trivial) 的：容易找到 TM M_1，满足 $L(M_1) = \emptyset$，以及 TM M_2，满足 $L(M_2) = \Sigma^*$。这样 $\langle M_1\rangle \notin L_3$，且 $\langle M_2\rangle \in L_3$。

由 Rice's Theorem，结论得证。

d) 这个语言（记作 L_4）是不可判定的，将 $\overline{A_{\mathsf{TM}}}$ 归约到 L_4。归约函数 f 满足 $f(\langle M, w\rangle) = \langle M_1, M_2\rangle$，其中 M_1 和 M_2 的定义如下：

$M_1 =$"对任意输入 x，

模拟 M 运行 w，若 M 接受 w，则 M_1 接受 x。"

$M_2 =$"对任意输入 x，拒绝。"

容易看出 f 是可计算的，$L(M_2) = \emptyset$，以及

1. $\langle M, w\rangle \in \overline{A_{\mathsf{TM}}} \Rightarrow w$ 被 M 拒绝或陷入死循环 $\Rightarrow L(M_1) = \emptyset \Rightarrow L(M_1) \subseteq L(M_2) \cup A = A$

   $\Rightarrow \langle M_1, M_2\rangle \in L_4$；

2. $\langle M, w\rangle \notin \overline{A_{\mathsf{TM}}} \Rightarrow w$ 被 M 接受 $\Rightarrow L(M_1) = \Sigma^* \Rightarrow L(M_1) \not\subseteq L(M_2) \cup A = A$

   $\Rightarrow \langle M_1, M_2\rangle \notin L_4$。

因为 $\overline{A_{\mathsf{TM}}}$ 是不可判定的，于是 L_4 也是不可判定的。

4. Consider the following generalization of Rice's Theorem: If P_2 is a non-trivial property of pairs of recognizable languages, then

$$A_{P_2} = \{\langle M, N\rangle \mid M \text{ and } N \text{ are Turing machines and } P_2(M, N) = \text{TRUE}\}$$

is undecidable.

证明: 将 A_{TM} 归约到 A_{P_2}。首先，根据非平凡性质，存在 M_1, N_1 满足 $P_2(M_1, N_1) = \text{TRUE}$，以及存在 M_2, N_2 满足 $f(M_2, N_2) = \text{FALSE}$。构造归约函数

$$f(\langle M, w\rangle) = \begin{cases} \langle M_1, N_1\rangle & \text{若 } M \text{ 接受 } w\text{，即 } \langle M, w\rangle \in A_{\mathsf{TM}} \\ \langle M_2, N_2\rangle & \text{其它} \end{cases}$$

显然 $\langle M, w\rangle \in A_{\mathsf{TM}} \Leftrightarrow f(\langle M, w\rangle) \in A_{P_2}$。由于 A_{TM} 是不可判定的，于是 A_{P_2} 也是不可判定的。

5. Define *TOTAL* to be the set of Turing machines that accept all strings of 0s and 1s:

$$TOTAL = \{\langle M\rangle \mid M \text{ is a Turing machine and } L(M) = \{\mathtt{0},\mathtt{1}\}^*\}$$

Using mapping reducibility to prove the following statements:

(a) *TOTAL* is not Turing-recognizable.

(b) The complement of *TOTAL* is not Turing-recognizable.

证明:

(a) 将 A_{TM} 归约到 $\overline{TOTAL}$，这与 $\overline{A_{\mathsf{TM}}} \leq_m TOTAL$ 等价。

归约函数 f 满足 $f(\langle M, w\rangle) = M'$，其中 M' 定义为:

$M' =$"对于任意输入 x， 　　模拟 M 运行 w 共 $\lvert x\rvert$ 步，如果此时 M 接受 w，拒绝；否则接受。"

显然 f 是可计算的，而且

1. $\langle M, w\rangle \in A_{\mathsf{TM}} \Rightarrow w$ 可以被 M 接受 $\Rightarrow$ 一定存在某 x 使得 M' 拒绝 $x \Rightarrow L(M') \neq \{\mathtt{0},\mathtt{1}\}^*$

   $\Rightarrow \langle M'\rangle \notin TOTAL \Rightarrow \langle M'\rangle \in \overline{TOTAL}$;

2. $\langle M, w\rangle \notin A_{\mathsf{TM}} \Rightarrow w$ 被 M 拒绝或陷入死循环 $\Rightarrow$ 任意输入 x 都会被 M' 接受 $\Rightarrow L(M') = \Sigma^*$

   $\Rightarrow \langle M'\rangle \in TOTAL \Rightarrow \langle M'\rangle \notin \overline{TOTAL}$。

于是 $A_{\mathsf{TM}} \leq_m \overline{TOTAL} \Rightarrow \overline{A_{\mathsf{TM}}} \leq_m TOTAL$。由于 $\overline{A_{\mathsf{TM}}}$ 不是 RE 的，于是 *TOTAL* 不是 RE 的。

(b) 将 A_{TM} 归约到 *TOTAL*，这与 $\overline{A_{\mathsf{TM}}} \leq_m \overline{TOTAL}$ 等价。

归约函数 f 满足 $f(\langle M, w\rangle) = M'$，其中 M' 定义为:

> $M' =$"对于任意输入 x，
>
> 模拟 M 运行 w，若 M 接受 w，接受 x。"

显然 f 是可计算的，而且

1. $\langle M, w\rangle \in A_{\mathsf{TM}} \Rightarrow w$ 可以被 M 接受 $\Rightarrow L(M') = \Sigma^* \Rightarrow \langle M'\rangle \in \textit{TOTAL}$;
2. $\langle M, w\rangle \notin A_{\mathsf{TM}} \Rightarrow w$ 在 M' 上被拒绝或陷入死循环 $\Rightarrow M'$ 不接受任何 $x \Rightarrow L(M') = \emptyset$

   $\Rightarrow \langle M'\rangle \notin \textit{TOTAL}$;

于是 $A_{\mathsf{TM}} \leq_m \textit{TOTAL} \Rightarrow \overline{A_{\mathsf{TM}}} \leq_m \overline{\textit{TOTAL}}$。由于 $\overline{A_{\mathsf{TM}}}$ 不是 RE 的，于是 $\overline{\textit{TOTAL}}$ 不是 RE 的。