Lecture 34

Rice's Theorem

Rice's theorem says that undecidability is the rule, not the exception. It is a very powerful theorem, subsuming many undecidability results that we have seen as special cases.

Theorem 34.1 **(Rice's theorem)** *Every nontrivial property of the r.e. sets is undecidable.*

Yes, you heard right: that's *every* nontrivial property of the r.e. sets. So as not to misinterpret this, let us clarify a few things.

First, fix a finite alphabet Σ. A *property of the r.e. sets* is a map

$$P : \{\text{r.e. subsets of } \Sigma^*\} \to \{\top, \bot\},$$

where $\top$ and $\bot$ represent truth and falsity, respectively. For example, the property of emptiness is represented by the map

$$P(A) = \begin{cases} \top & \text{if } A = \varnothing, \\ \bot & \text{if } A \neq \varnothing. \end{cases}$$

To ask whether such a property P is decidable, the set has to be presented in a finite form suitable for input to a TM. We assume that r.e. sets are presented by TMs that accept them. But keep in mind that the property is a property of *sets*, not of Turing machines; thus it must be true or false independent of the particular TM chosen to represent the set.

Here are some other examples of properties of r.e. sets: $L(M)$ is finite; $L(M)$ is regular; $L(M)$ is a CFL; M accepts 101001 (i.e., $101001 \in L(M)$); $L(M) = \Sigma^*$. Each of these properties is a property of the set accepted by the Turing machine.

Here are some examples of properties of Turing machines that are *not* properties of r.e. sets: M has at least 481 states; M halts on all inputs; M rejects 101001; there exists a smaller machine equivalent to M. These are not properties of sets, because in each case one can give two TMs that accept the same set, one of which satisfies the property and the other of which doesn't.

For Rice's theorem to apply, the property also has to be *nontrivial.* This just means that the property is neither universally true nor universally false; that is, there must be at least one r.e. set that satisfies the property and at least one that does not. There are only two trivial properties, and they are both trivially decidable.

Proof of Rice's theorem. Let P be a nontrivial property of the r.e. sets. Assume without loss of generality that $P(\varnothing) = \bot$ (the argument is symmetric if $P(\varnothing) = \top$). Since P is nontrivial, there must exist an r.e. set A such that $P(A) = \top$. Let K be a TM accepting A.

We reduce HP to the set $\{M \mid P(L(M)) = \top\}$, thereby showing that the latter is undecidable (Theorem 33.3(ii)). Given $M\#x$, construct a machine $M' = \sigma(M\#x)$ that on input y

(i) saves y on a separate track someplace;

(ii) writes x on its tape (x is hard-wired in the finite control of M');

(iii) runs M on input x (a description of M is also hard-wired in the finite control of M');

(iv) if M halts on x, M' runs K on y and accepts if K accepts.

Now either M halts on x or not. If M does not halt on x, then the simulation in (iii) will never halt, and the input y of M' will not be accepted. This is true for every y, so in this case $L(M') = \varnothing$. On the other hand, if M does halt on x, then M' always reaches step (iv), and the original input y of M' is accepted iff y is accepted by K; that is, if $y \in A$. Thus

$$\begin{array}{rcccl} M \text{ halts on } x & \Rightarrow & L(M') = A & \Rightarrow & P(L(M')) = P(A) = \top, \\ M \text{ does not halt on } x & \Rightarrow & L(M') = \varnothing & \Rightarrow & P(L(M')) = P(\varnothing) = \bot. \end{array}$$

This constitutes a reduction from HP to the set $\{M \mid P(L(M)) = \top\}$. Since HP is not recursive, by Theorem 33.3, neither is the latter set; that is, it is undecidable whether $L(M)$ satisfies P. □

Rice's Theorem, Part II

A property $P : \{\text{r.e. sets}\} \to \{\top, \bot\}$ of the r.e. sets is called *monotone* if for all r.e. sets A and B, if $A \subseteq B$, then $P(A) \le P(B)$. Here $\le$ means less than or equal to in the order $\bot \le \top$. In other words, P is *monotone* if whenever a set has the property, then all supersets of that set have it as well. For example, the properties "$L(M)$ is infinite" and "$L(M) = \Sigma^*$" are monotone but "$L(M)$ is finite" and "$L(M) = \varnothing$" are not.

Theorem 34.2 **(Rice's theorem, part II)** *No nonmonotone property of the r.e. sets is semidecidable. In other words, if P is a nonmonotone property of the r.e. sets, then the set $T_P = \{M \mid P(L(M)) = \top\}$ is not r.e.*

Proof. Since P is nonmonotone, there exist TMs M_0 and M_1 such that $L(M_0) \subseteq L(M_1)$, $P(M_0) = \top$, and $P(M_1) = \bot$.

We want to reduce $\sim$HP to T_P, or equivalently, HP to $\sim T_P = \{M \mid P(L(M)) = \bot\}$. Since $\sim$HP is not r.e., neither will be T_P. Given $M\#x$, we want to show how to construct a machine M' such that $P(M') = \bot$ iff M halts on x. Let M' be a machine that does the following on input y:

(i) writes its input y on the top and middle tracks of its tape;

(ii) writes x on the bottom track (it has x hard-wired in its finite control);

(iii) simulates M_0 on input y on the top track, M_1 on input y on the middle track, and M on input x on the bottom track in a round-robin fashion; that is, it simulates one step of each of the three machines, then another step, and so on (descriptions of M_0, M_1, and M are all hard-wired in the finite control of M');

(iv) accepts its input y if either of the following two events occurs:

   (a) M_0 accepts y, or

   (b) M_1 accepts y and M halts on x.

Either M halts on x or not, independent of the input y to M'. If M does not halt on x, then event (b) in step (iv) will never occur, so M' will accept y iff event (a) occurs, thus in this case $L(M') = L(M_0)$. On the other hand, if M does halt on x, then y will be accepted iff it is accepted by either M_0 or M_1; that is, if $y \in L(M_0) \cup L(M_1)$. Since $L(M_0) \subseteq L(M_1)$, this is equivalent to saying that $y \in L(M_1)$, thus in this case $L(M') = L(M_1)$. We have shown

$$\begin{aligned} M \text{ halts on } x &\Rightarrow L(M') = L(M_1) \\ &\Rightarrow P(L(M')) = P(L(M_1)) = \bot, \end{aligned}$$

$$\begin{aligned} M \text{ does not halt on } x &\Rightarrow L(M') = L(M_0) \\ &\Rightarrow P(L(M')) = P(L(M_0)) = \top. \end{aligned}$$

The construction of M' from M and x constitutes a reduction from $\sim$HP to the set $T_P = \{M \mid P(L(M)) = \top\}$. By Theorem 33.3(i), the latter set is not r.e. □

Historical Notes

Rice's theorem was proved by H. G. Rice [104, 105].