

Практика №2 по курсу «Теория алгоритмов»
«Машина Тьюринга»
Группы ФТ-301, ФТ-302

1 Распознаватели и работа над ошибками

На данной практике мы будем рассматривать МТ-распознаватели, то есть МТ с двумя терминальными состояниями q_{accept} и q_{reject} . Напомним ещё раз, что такие МТ описываются кортежем $\langle \Gamma, Q, \delta, q_{start}, q_{accept}, q_{reject} \rangle$.

В конспекте прошлой практики было написано, что **грубо говоря** можно считать, что МТ распознаватель вычисляет некоторую функцию $f : \Sigma^* \mapsto \{0, 1, \perp\}$, где $\Sigma = \Gamma \setminus \{\lambda\}$, а \perp — специальный символ для обозначения заиклившейся или сломанной МТ. Данная интерпретация неверна, так как из неё ускользает аспект неразличимости зависшей МТ от долго работающей МТ. Поэтому корректнее считать, что МТ-распознаватель вычисляет некоторую частичную функцию $f : D \mapsto \{0, 1\}$, где $D \subseteq \Sigma^*$ — область определения частичной функции f . Различие этих определений заключается в том, что в первом случае МТ якобы «знает» о своей области определения, а во втором — нет.

В качестве небольшого примера для улучшения понимания «заикливающих» МТ можно рассмотреть следующее плохое определение МТ:

Определение 1. Пусть на вход МТ подается описание полинома p над множеством целых чисел с переменными x_1, x_2, \dots, x_k . МТ работает следующим образом:

- Перебрать все возможные целочисленные комбинации переменных x_1, x_2, \dots, x_k ;
- Вычислить p на каждом из этих наборов параметров;
- Если для некоторого набора параметров полином равен 0, то перейти в состояние q_{accept} , иначе — в q_{reject} .

Проблема данного определения МТ состоит в том, что МТ никогда не сможет попасть в состояние q_{reject} , так как вычислить значение полинома во всех целочисленных точках невозможно (их бесконечное число).

2 Разрешимые и перечислимые языки

Для машины Тьюринга M обозначим за $RCG(M)$ множество строк, на которых МТ заканчивает свою работу в состоянии q_{accept} . Будем называть множество (язык) L *перечислимым*, если существует МТ M такая, что $RCG(M) = L$.

Будем называть МТ M *решающей*, если для любого входа $w \in \Sigma^*$ данная МТ заканчивает свою работу в одном из состояний q_{accept} или q_{reject} , то есть не ломается и не заикливается. В таком случае будем называть множество (язык) L *разрешимым*, если существует решающая МТ M такая, что $RCG(M) = L$.

Пример 1. Множество полиномов L с целочисленным корнем, определённое выше, очевидно является перечислимым (в качестве перечисляющей МТ подойдет машина, перебирающая все значения параметров в правильном порядке, например в порядке возрастания суммы $\sum |x_i|$)

$$L = \{ \langle p \rangle \mid p - \text{многочлен с целочисленным корнем над переменными } x_1, x_2, \dots, x_k \}$$

Вопрос о разрешимости этого языка не так тривиален и был внесён в список 23 проблем Гильберта под номером 10. Доказательство алгоритмической неразрешимости произвольного диофантова уравнения (то есть неразрешимости языка L) была доказана только в 1970 году.

Пример 2. Рассмотрим несколько похожий язык:

$$L_{DFA} = \{ \langle A, B \rangle \mid L(A) = L(B), \text{ где } A, B - \text{конечные автоматы над бинарным алфавитом, а } L(D) - \text{регулярный язык, задающийся автоматом } D \}$$

Легко заметить, что данный язык является перечислимым, так как достаточно просто перебрать все бинарные строки в порядке возрастания длины и синхронно для каждой строки проверять принадлежность каждому из автоматов A и B .

Однако, в отличие от примера с полиномами, данный язык также является разрешимым. Для этого достаточно доказать следующее утверждение:

Замечание 3. Обозначим за n, m количество состояний в автоматах A и B соответственно. Тогда, если существует слово $w \in \{0, 1\}^*$ такое, что $w \in L(A), w \notin L(B)$ и $|w| \geq n \times m$, то существует слово $w', |w'| < |w|$ такое, что $w' \in L(A), w' \notin L(B)$.

Доказательство. Определим значение функцию $states(w[1..i]) = \langle A(w[1..i]), B(w[1..i]) \rangle$ как кортеж из пары состояний, в которых находятся автоматы A и B после прочтения строки $w[1..i]$. Заметим, что всего различных пар состояний $n \times m$, а для строки w существует $|w| + 1 > n \times m$ значений функции $states$ (учитывая начальное состояние для пустой строки). Значит по принципу Дирихле найдется пара позиций $i \neq j$ таких, что $states(w[1..i]) = states(w[1..j])$, а значит можно рассмотреть строку $w' = w[1..i] + w[j + 1..|w|]$, длина которой строго меньше длины $|w|$ и при этом $states(w') = states(w)$, а значит $w' \in L(A)$ и $w' \notin L(B)$. \square

Таким образом, если языки различаются, то существует строка w короче $n \times m$ символов, которая входит в один язык и не входит в другой.

Тогда чтобы построить решающую МТ для языка L_{DFA} достаточно перебрать все бинарные строки длиной до $n \times m$ символов.

На практиках также предлагалось решение, которое минимизирует пару исходных автоматов, а дальше сравнивает графы на изоморфность. Данный подход также верный, однако для доказательства его корректности необходимо показать, что для фиксированного регулярного языка все минимальные ДКА будут изоморфны друг другу. Данный факт следует из теоремы Майхилла-Нероуда (Myhill–Nerode theorem).

3 Неперечеслимые языки

Мы знаем, что язык $L_{TM} = \{\langle M, w \rangle \mid M \text{ останавливается на } w\}$ является неразрешимым, но перечислимым языком. Постараемся найти неперечислимый язык, т.е. язык L такой, что не существует МТ $M : RCG(M) = L$. Для этого докажем следующее утверждение:

Лемма 4. (теорема Поста) Язык L является разрешимым тогда и только тогда, когда L и \bar{L} — перечислимые языки

Доказательство. Ясно, что из разрешимости L следует перечислимость L и \bar{L} .

Доказательство в обратную сторону также достаточно простое — так как для L и \bar{L} существуют распознаватели, то достаточно запустить их параллельно, чтобы построить решатель для языка L . \square

Таким образом из леммы доказанной выше сразу следует, что $\overline{L_{TM}}$ является неперечислимым языком.

4 Операции над разрешимыми и перечислимыми языками

Лемма 5. Для любого разрешимого языка L верно, что $\bar{L} = U \setminus L$ также является разрешимым языком

Доказательство. Для L существует решающая МТ $M = \langle \Gamma, Q, \delta, q_{start}, q_{accept}, q_{reject} \rangle$. Несложно показать, что $M' = \langle \Gamma, Q, \delta, q_{start}, q_{reject}, q_{accept} \rangle$ является решающей МТ для языка \bar{L} . \square

Вопросы на дом:

- Правда ли, что для любого конечного набора разрешимых языков L_1, L_2, \dots, L_k верно, что $L = \cup_i L_i$ также является разрешимым языком?
- Правда ли, что разрешимые языки замкнуты относительно операции гомоморфизма $h : \Sigma \mapsto \Delta^*$?
Напомним, что $h(L) = \{h(w) \mid w \in L\}$, а $h(w_1 w_2 \dots w_n) = h(w_1) h(w_2) \dots h(w_n)$
- Правда ли, что для любого перечислимого языка L и некоторого гомоморфизма h верно, что $h(L)$ также является перечислимым?