

Детерминизация конечного автомата

Кузнецов А.Д.

Оглавление

1	Краткая теоретическая часть	2
1.1	Автомат	2
1.2	Теорема Клини	3
2	Постановка задачи	3
3	Алгоритм детерминизации автомата	4
3.1	Удаление λ -переходов	4
3.2	Детерминизация КНА	4
3.3	Оценка сложности алгоритма	4
3.4	Обоснование корректности работы алгоритма	4
4	Реализация алгоритма детерминизации автомата	4
4.1	Кодирование автоматов (реализация классов)	4
4.2	Реализация алгоритма удаления λ -переходов	4
4.3	Реализация алгоритма детерминизации КНА	4
4.4	Уязвимые для критики места	4
5	Тестирование	4
5.1	Unit-тестирование	4
5.2	Умное тестирование	4
6	Использование алгоритма	4
6.1	Формат файла-автомата (.fsa)	4
6.2	Компиляция и запуск основной программы	4
7	Заключение	4

1 Краткая теоретическая часть

Ниже приводятся определения и утверждения, которые будут использованы для дальнейших пояснений к реализации алгоритма детерминизации.

1.1 Автомат

Конечный недетерминированный автомат (КНА) M — это кортеж вида

$$M = (A, Q, q_0, F, \delta),$$

где

- $A = \{a_0, a_1, \dots, a_{m-1}\}$ — входной алфавит, т.е. множество символов, причем конечное: $|A| = m$;
- $Q = \{q_0, q_1, \dots, q_{n-1}\}$ — множество состояний автомата, тоже конечное: $|Q| = n$;
- q_0 — начальное состояние автомата, то есть $q_0 \in Q$;
- $F \subset Q$ — выходные состояния автомата;
- $\delta : Q \times A \rightarrow 2^Q$ — функция переходов автомата.

Если в автомате разрешены переходы по пустому символу, то входной алфавит дополняется фиктивным символом λ (в иностранной литературе его чаще обозначают ε). В таком случае автомат называют λ -КНА (ε -FSA). Входной алфавит при этом будем обозначать так же, то есть если речь идет о λ -КНА, то подразумевается, что $A \leftarrow A \cup \{\lambda\}$, а мощность $|A| = m + 1$.

Конечный детерминированный автомат (КДА) M — это такой КНА, где функция переходов δ выглядит так:

$$\delta : Q \times A \rightarrow 2^Q,$$

т.е. из любого состояния по любой букве возможен переход в точности в одно состояние — это и обеспечивает детерминированность работы такого автомата.

1.2 Теорема Клини

Пусть $A = \{a_0, \dots, a_{n-1}\}$ - произвольный алфавит. Язык $L \subseteq A^*$ является элементом полукольца регулярных языков $R(A)$ в алфавите A тогда и только тогда, когда он допускается некоторым конечным автоматом.

2 Постановка задачи

Теперь сформулируем задачу в вышеуказанных терминах:

Реализовать алгоритм преобразования λ -КНА M в КНА \hat{M} так, чтобы распознаваемые ими языки совпадали, т.е. $L(M) = L(\hat{M})$.

3 Алгоритм детерминизации автомата

3.1 Удаление λ -переходов

3.2 Детерминизация КНА

3.3 Оценка сложности алгоритма

3.4 Обоснование корректности работы алгоритма

4 Реализация алгоритма детерминизации автомата

4.1 Кодирование автоматов (реализация классов)

4.2 Реализация алгоритма удаления λ -переходов

4.3 Реализация алгоритма детерминизации КНА

4.4 Уязвимые для критики места

5 Тестирование

5.1 Unit-тестирование

5.2 Умное тестирование

6 Использование алгоритма

6.1 Формат файла-автомата (.fsa)

6.2 Компиляция и запуск основной программы

7 Заключение

~~Реализовано все круто, добавить нечего, почтаще бы так писали код.~~