Теория автоматов и формальных языков

Магазинные автоматы

Автор: Екатерина Вербицкая

Санкт-Петербургский государственный электротехнический университет «ЛЭТИ»

30 ноября 2021

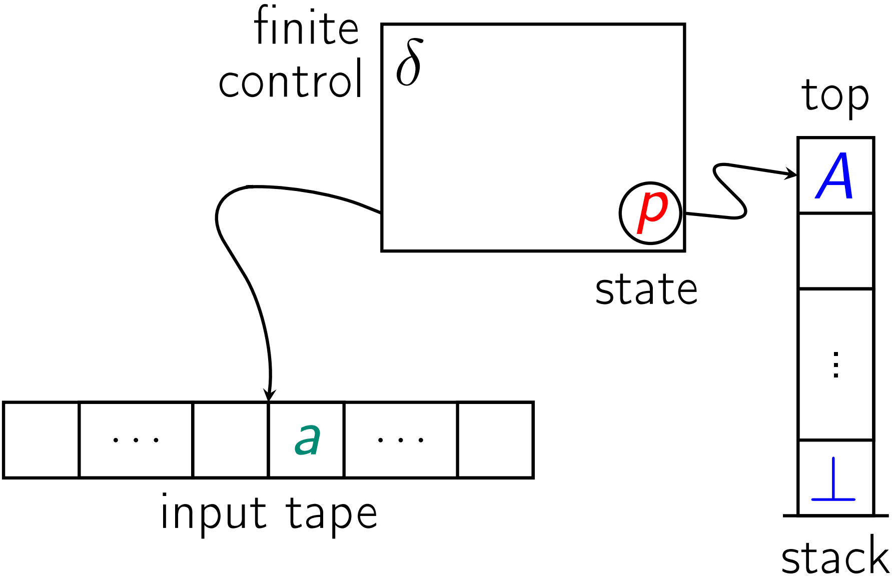
В предыдущей серии

- Регулярные языки распознаются с помощью конечных автоматов
- Разные алгоритмы синтаксического анализа для контекстно-свободных языков
  - CYK
  - Рекурсивный спуск
  - LL(1)
  - LR(0), SLR(1), CLR(1), LALR(1)
- Есть ли универсальный распознаватель для КС-языков?

TLDR

- Произвольный КС язык можно распознать при помощи магазинного автомата (он же автомат с магазинной памятью, он же pushdown automata, он же pda)
- Магазинный автомат по сути — автомат со стеком
- Детерминированные магазинные автоматы могут распознавать только детерминированные КС языки
- Недетерминированные магазинные автоматы могут распознавать произольные КС языки

Что такое магазинный автомат



Что такое магазинный автомат: неформально

- Автомат, переходы которого осуществляются по входному символу, текущему состоянию и символу на вершине стека
  - У конечного автомата не было стека
- Никакие состояния стека, кроме вершины, не доступны
- Во время перехода может изменяться стек
  - Положить что-то на стек (push)
  - Снять верхушку со стека (pop)
- А может и не изменяться
  - Магазинный автомат может вообще игнорировать стек
  - Или стек может не изменяться, хоть значение оттуда и читается
- Итого: по тройке (входной символ, состояние, символ на вершине стека) получается новое состояние, и модифицируется (или нет) стек

Формальное определение

Недетерминированный магазинный автомат это $(Q, \Sigma, \Gamma, \delta, q_0, \perp, F)$

- Q — конечное множество состояний
- Σ — конечное множество символов, входной алфавит
- Γ — конечное множество символов, стековый алфавит
- $\delta : Q \times (\Sigma \cup \varepsilon) \times \Gamma \to 2^{Q \times \Gamma^*}$ — функция переходов
- $q_0 \in Q$ — стартовое состояние
- $\perp \in \Gamma$ — начальный элемент стека
- $F \subseteq Q$ — множество принимающих (конечных) состояний

Отношение переходов

$$(q, \alpha) \in \delta(p, a, A) \text{ означает}$$

- Если магазинный автомат находится в состоянии $p \in Q$,
- на вершине стека находится $A \in \Gamma$,
- а со входа читается символ $a \in \Sigma \cup \varepsilon$,

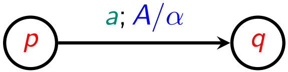
- то изменяем состояние на $q \in Q$,
- снимаем со стека символ A, записываем на стек строку $\alpha \in \Gamma^*$

- $\Sigma \cup \varepsilon$ сигнализирует о том, что вход можно и не читать

$$p \xrightarrow{a;\, A/\alpha} q$$

Семантика магазинного автомата

- Мгновенное описание МА: $(p, \omega, \beta) \in Q \times \Sigma^* \times \Gamma^*$
  - p — текущее состояние автомата
  - ω — непрочитанный фрагмент входного потока
  - β — содержимое стека (верхушка записана первой)
- Отношение $\vdash$ на мгновенных описаниях (шаг)
  - Для каждого $(q, \alpha) \in \delta(p, a, A)$, верно $(p, a\omega, A\gamma) \vdash (q, \omega, \alpha\gamma)$ для произвольных $\omega \in \Sigma^*, \gamma \in \Gamma^*$
- Шаг не определен, если стек пуст

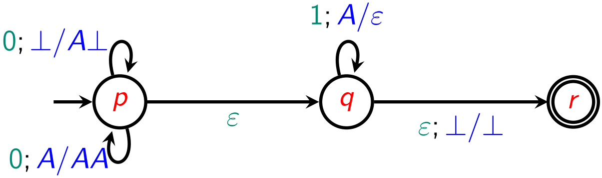


Семантика магазинного автомата: вычисление

- Вычисление — последовательность шагов
- Начальное мгновенное описание $(q_0, \omega, \perp)$
- Выбирается любой из подходящих шагов
- Если какой-нибудь выбор приведет к успеху, значит, строка распознается

- Два варианта окончания работы
  - По достижении конечного состояния
    - $L(M) = \{\omega \in \Sigma^* \mid (q_0, \omega, \perp) \vdash^* (f, \varepsilon, \gamma), f \in F, \gamma \in \Gamma^*\}$
  - По опустошении стека
    - $N(M) = \{\omega \in \Sigma^* \mid (q_0, \omega, \perp)) \vdash^* (q, \varepsilon, \varepsilon), q \in Q\}$
  - Эти варианты эквивалентны: по автомату, завершающемуся по первой схеме, можно посмотроить автомат, завершающийся по второй схеме, и наоборот
- $\vdash^*$ — транзитивно рефлексивное замыкание отношения $\vdash$

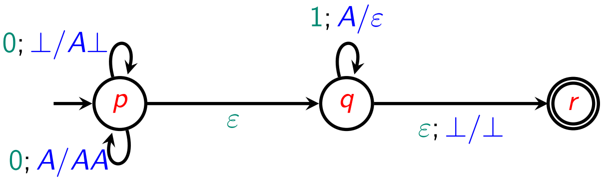
Пример: язык $\{0^n 1^n \mid n \geq 0\}$ (конечное состояние)



Вычисление на строке 0011:

- $(p, 0011, \perp) \vdash (q, 0011, \perp) \vdash (r, 0011, \perp)$ — провал
- $(p, 0011, \perp) \vdash (p, 011, A\perp) \vdash (q, 011, A\perp)$ — провал
- $(p, 0011, \perp) \vdash (p, 011, A\perp) \vdash (p, 11, AA\perp) \vdash (q, 11, AA\perp) \vdash (q, 1, A\perp) \vdash (q, \varepsilon, \perp) \vdash (r, \varepsilon, \perp)$ — успех (по принимающему состоянию)

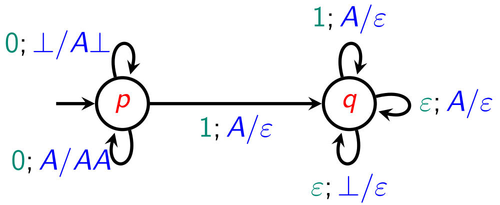
Пример: язык $\{0^n1^n \mid n \geq 0\}$ (конечное состояние)



Вычисление на строке 00111:

- $(p, 00111, \bot) \vdash (q, 00111, \bot) \vdash (r, 00111, \bot)$ — провал
- $(p, 00111, \bot) \vdash (p, 0111, A\bot) \vdash (q, 0111, A\bot)$ — провал
- $(p, 00111, \bot) \vdash (p, 0111, A\bot) \vdash (p, 111, AA\bot) \vdash (q, 111, AA\bot) \vdash (q, 11, A\bot) \vdash (q, 1, \bot) \vdash (r, 1, \bot)$ — провал

Пример: язык $\{0^n1^m \mid 1 \leq m \leq n\}$ (опустошение стека)



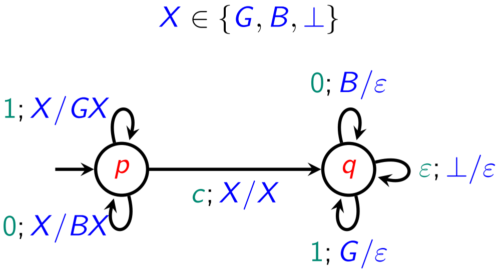
Вычисление на строке 00111:

- $(p, 00111, \bot) \vdash (q, 00111, \bot)$ — провал
- $(p, 00111, \bot) \vdash (p, 0111, A\bot) \vdash (p, 111, AA\bot) \vdash (q, 11, A\bot) \vdash (q, 1, \bot)$ — провал

Вычисление на строке 001:

- $(p, 001, \bot) \vdash (p, 01, A\bot) \vdash (p, 1, AA\bot) \vdash (q, \varepsilon, A\bot) \vdash (q, \varepsilon, \bot) \vdash (q, \varepsilon, \varepsilon)$ — успех

Пример: язык $\{\omega c \omega^R \mid \omega \in \{0,1\}^*\}$ (опустошение стека)



Вычисление на строке $110c011$:

- $(p, 110c011, \perp) \vdash (p, 10c011, G\perp) \vdash (p, 0c011, GG\perp) \vdash (p, c011, BGG\perp) \vdash (q, 011, BGG\perp) \vdash (q, 11, GG\perp) \vdash (q, 1, G\perp) \vdash (q, \varepsilon, \perp) \vdash (q, \varepsilon, \varepsilon)$ — успех

Формальное определение ДМА

Детерминированный магазинный автомат это $(Q, \Sigma, \Gamma, \delta, q_0, \perp, F)$

- Q — конечное множество состояний
- Σ — конечное множество символов, входной алфавит
- Γ — конечное множество символов, стековый алфавит
- $\delta : Q \times (\Sigma \cup \varepsilon) \times \Gamma \to Q \times \Gamma^*$ — функция переходов
  - $\forall q \in Q, Z \in \Gamma$: если есть эпсилон-переход $\delta(q, \varepsilon, Z)$ то нет переходов по терминалам $\delta(q, a, Z)$
- $q_0 \in Q$ — стартовое состояние
- $\perp \in \Gamma$ — начальный элемент стека
- $F \subseteq Q$ — множество принимающих (конечных) состояний

Детерминированные магазинные автоматы vs недетерминированные

- В общем случае одной входной строке может соответствовать несколько вычислений
  - Некоторые из них могут завершаться в принимающих состояниях
- Если существует хотя бы одно вычисление, завершающееся в принимающем состоянии, строка принадлежит языку
- Если для каждой строки существует ровно одно вычисление в магазинном автомате, то он является *детерминированным*
  - Соответствующий язык является *детерминированным КС языком*
- Детерминированный магазинный автомат является частным случаем недетерминированного, поэтому детерминированные КС языки — строгое подмножество контекстно-свободных

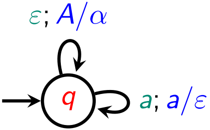
Неэквивалентность двух видов приема для детерминированных магазинных автоматов

Беспрефиксный язык — язык, в котором никакое слово не является префиксом другого

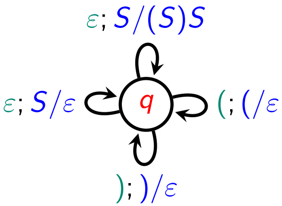
- Прием языка детерминированным магазинным автоматом по пустому стеку и по допускающему состоянию эквивалентно только для беспрефиксных языков
- Рассмотрим слово $\omega = \alpha\beta : \alpha, \beta \in \Sigma^*, \omega, \alpha \in L$, где $L \subseteq \Sigma^*$
- При попытке распознать слово ω ДМП завершит свою работу, как только прочитает α
- ω никогда не будет принята
- Можно построить ДМП, принимающий по допускающему состоянию, который допускает префиксный язык

Построение магазинного автомата по КС-грамматике

- Интуиция:
  - Для каждого нетерминала: заменяем его на стеке на правую часть правила
  - Для каждого терминала: считываем со входа этот терминал и снимаем со стека
- Построение:
  - Для каждого правила $A \to \alpha$ добавляем (q, α) в $\delta(q, \varepsilon, A)$
  - Для каждого терминала a добавляем (q, ε) в $\delta(q, a, a)$
- Относительно бесполезный автомат: как найти правильное вычисление?



Магазинный автомат по грамматике $S \to (S)S \mid \varepsilon$

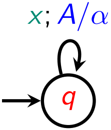


Вычисление на строке (())():

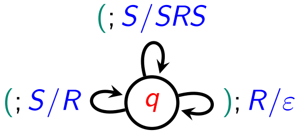
- $(q, (())(), S) \vdash (q, (())(), (S)S) \vdash (q, ())(), S)S) \vdash$
  $(q, ())(), (S)S)S) \vdash (q,))(), S)S)S) \vdash (q,))(),)S)S) \vdash (q,)(), S)S) \vdash$
  $(q,)(),)S) \vdash (q, (), S) \vdash (q, (), (S)S) \vdash (q,), S)S) \vdash (q,),)S) \vdash$
  $(q, \varepsilon, S) \vdash (q, \varepsilon, \varepsilon)$ — успех

Построение магазинного автомата по КС-грамматике в нормальной форме Грейбах

- Нормальная форма Грейбах: все правила имеют вид $A \to x\alpha$
  - $A \in V_N$
  - $x \in V_T$
  - $\alpha \in V_N^*$
- Построение магазинного автомата:
  - Для каждого правила $A \to x\alpha$ добавляем (q, α) в $\delta(q, x, A)$
- Автоматы получаются проще
- Не сильно более полезный автомат: нужны алгоритмы синтаксического анализа



Магазинный автомат по грамматике в НФГ
$S \to (SRS \mid (R$
$R \to \,)$



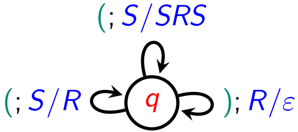
Вычисление на строке (())():

- $(q, (())(), S) \vdash (q, ())(), SRS) \vdash (q,))(), RRS) \vdash (q,)(), RS) \vdash$ $(q, (), S) \vdash (q,), R) \vdash (q, \varepsilon, \varepsilon)$ — успех

Построение КС грамматики по магазинному автомату

- По магазинному автомату $(Q, \Sigma, \Gamma, \delta, q_0, \perp, F)$ строим грамматику (V_N, V_T, P, S)
- $V_N = \{S\} \cup \{[qAp] \mid p, q \in Q, A \in \Gamma\}$
- $V_T = \Sigma$
- $P = \{S \to [q_0 \perp q] \mid q \in Q\} \cup$
  $\{[qAq_{m+1}] \to a[q_1 B_1 q_2][q_2 B_2 q_3] \dots [q_m B_m q_{m+1}] \mid$
  $q_i \in Q; a \in \Sigma \cup \{\varepsilon\}; A, B_j \in \Gamma; (q_1, B_1 B_2 \dots B_m) \in \delta(q, a, A)\} \cup$
  $\{[qAp] \to a \mid (p, \varepsilon) \in \delta(q, a, A)\}$

Пример построения КС грамматики по магазинному автомату

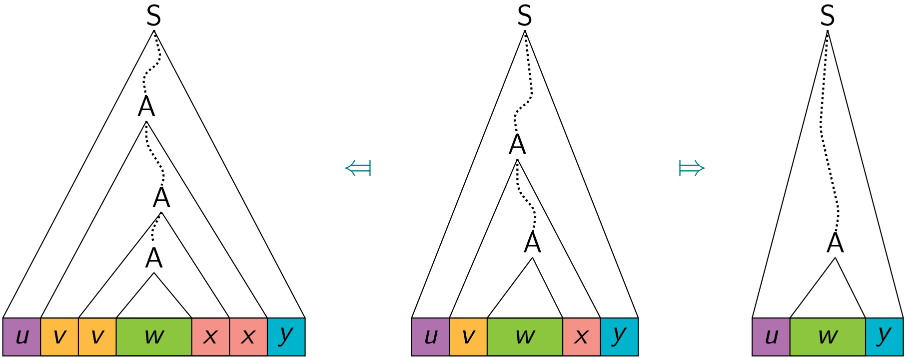


$$
\begin{array}{rcl}
[qSq] & \rightarrow & (\ [qRq] \\
 & \mid & (\ [qSq]\ [qRq]\ [qSq] \\
[qRq] & \rightarrow &)
\end{array}
$$

Лемма о накачке для КС языков

Лемма

Если язык L является контекстно-свободным, то
$\exists p \geq 1 : \forall s \in L : |s| \geq p$ *можно разбить на подстроки*
$s = uvwxy : |vwx| \leq p, |vx| \geq 1$ *и*
$\forall n \geq 0 : \; uv^n wx^n y \in L$



Лемма о накачке для КС языков: пример

Язык $L = \{a^n b^n c^n\}$
Предполагаем, что он КС, тогда по Лемме существует p...
Рассмотрим слово $a^p b^p c^p = uvwxy, |vwx| \leq p, |vx| \geq 1$

- $vwx = a^j, j \leq p$
- $vwx = a^j b^k, j + k \leq p$
- $vwx = b^j, j \leq p$
- $vwx = b^j c^k, j + k \leq p$
- $vwx = c^j, j \leq p$

Строка $uv^i wx^i y$ не содержит одинаковое количество букв для всех i.
Например, рассмотреть $i = 2$. Получили противоречие — успех