О выразительных свойствах языка образцов Рефала

Антонина Непейвода

ИПС РАН

Рабочее совещание МГТУ им. Н. Э. Баумана и ИПС РАН по Рефалу
5 июня 2018

Определения

$\mathcal{V}_{\mathcal{T}}$ — множество переменных типа $\mathcal{T}$, $\mathcal{V} = \bigcup^{\mathcal{T}} \mathcal{V}_{\mathcal{T}}$. Алфавит Σ — множество констант. *По умолчанию считаем, что это множество не ограничено!*

Плоский образец P — строка в алфавите $\mathcal{V} \cup \Sigma$. Далее подразумеваем, что все образцы плоские. Подстановка σ — морфизм из $(\mathcal{V} \cup \Sigma)^*$ в Σ^*, сохраняющий константы (т.е. для всех $\mathbf{A} \in \Sigma$ $\sigma(\mathbf{A}) = \mathbf{A}$).

Определение

Язык $\mathscr{L}(\mathrm{P})$, *распознаваемый образцом* P — *множество элементов* $\Phi \in \Sigma^*$, *для которых существует подстановка* σ: $\sigma(\mathrm{P}) = \Phi$. *Образец* P_1 *сводится к образцу* P_2, *если* $\mathscr{L}(\mathrm{P}_1) \subseteq \mathscr{L}(\mathrm{P}_2)$.

- Язык, распознаваемый образцом-строкой $\mathrm{P} \in \Sigma^*$, есть $\{\mathrm{P}\}$.
- Язык, распознаваемый образцом $\mathrm{P} = \mathrm{e}.1\,\mathrm{e}.2 \dots \mathrm{e}.n$, есть все множество Σ^*.

Язык, распознаваемый набором образцов

$$
\begin{aligned}
&\mathtt{F}\,\{\\
&\qquad \mathsf{P}_1 \quad = u_1;\\
&\qquad \dots\\
&\qquad \mathsf{P}_n \quad = u_n;\\
&\}
\end{aligned}
$$

$u_i \in \{\mathbf{T}, \mathbf{F}\}$. Множество строк, распознаваемых образцом P_i внутри такого определения, есть $\mathscr{L}(\mathsf{P}_i) \setminus \bigcup_{j=1}^{i-1} \mathscr{L}(\mathsf{P}_j)$.

Определение

Будем называть языком $\mathscr{L}(\mathtt{F})$, распознаваемым $\mathtt{F}$, множество слов $\Phi \in \Sigma^*$, для которых $<\mathtt{F}\ \Phi> = \mathbf{T}$.

- Заменой всех u_i на их отрицания в определении $\mathtt{F}$ получим функцию, определенную на том же множестве, что исходная, и определяющую язык, являющийся дополнением к $\mathscr{L}(\mathtt{F})$.

Языки образцов без повторных e-переменных

- $\mathsf{P} \in (\mathcal{V}_\mathsf{e} \cup \Sigma)^* \Rightarrow \mathsf{P}$ распознает регулярные языки.
- Допускаем s-переменные — регулярность пропадает, если Σ неограничен. Для любого конечного заранее определенного Σ — регулярны.
- Без s-переменных — сводимость образца P_1 к образцу P_2 эквивалентна существованию сводящей подстановки.
- С s-переменными — не обязательно. $\mathsf{P}_1 = \mathsf{s}.1\,\mathsf{e}.x$, $\mathsf{P}_2 = \mathsf{e}.x\,\mathsf{s}.1$ эквивалентны, но сводящей подстановки не существует.

Языки, распознаваемые функциями без повторных е-переменных

Вопрос: Верно ли, что в случае без s-переменных только регулярные языки распознаются функцией, имеющей образцы без повторных е-переменных?

Еще вопрос: Верно ли, что каждый регулярный язык может быть распознан некоторой такой функцией?

Языки, распознаваемые функциями без повторных e-переменных

Вопрос: Верно ли, что в случае без s-переменных только регулярные языки распознаются функцией, имеющей образцы без повторных e-переменных?

Да, верно, в силу замкнутости регулярных языков относительно операций над множествами. Пусть функция $\mathtt{F}$ задается последовательностью правил $\mathsf{P}_i = u_i$. Для всех i таких, что $u_i = \mathbf{T}$, опишем распознаваемый i-м правилом язык.

$$\mathscr{L}_{\mathtt{F}}(\mathsf{P}_i) = \mathscr{L}(\mathsf{P}_i) \setminus \bigcup_{j<i} \mathscr{L}(\mathsf{P}_j).$$

Язык $\mathscr{L}(\mathtt{F})$ есть объединение этих языков.

Языки, распознаваемые функциями без повторных e-переменных

Еще вопрос: Верно ли, что каждый регулярный язык может быть распознан некоторой такой функцией?

Неверно. Пусть язык $(\mathbf{A}\,\mathbf{A})^*$ распознается функцией $\mathtt{F}$ указанного вида.

- Хотя бы одна строка функции $\mathtt{F}$ имеет вид $\mathsf{P}_i = \mathbf{T}$, $\mathsf{P}_i \in (\{\mathbf{A}\} \cup \mathcal{V})^+$, причем P_i содержит хотя бы одну e−переменную (иначе $\mathscr{L}(\mathtt{F})$ содержит лишь конечное число слов из $(\mathbf{A}\,\mathbf{A})^*$).
- Ни одна строка функции $\mathtt{F}$ выше i-ой не может иметь вид $\mathsf{P}_j = \mathbf{F}$, $\mathsf{P}_j \in (\{\mathbf{A}\} \cup \mathcal{V})^+$, если P_j содержит хотя бы одну e−переменную (иначе, подставив в e-переменные образца P_j достаточно длинные строки из букв $\mathbf{A}$, суммарная длина которых имеет ту же четность, что суммарное количество s-переменных и букв $\mathbf{A}$ в P_j, получим, что строка из языка $(\mathbf{A}\,\mathbf{A}^*)$ не входит в $\mathscr{L}(\mathtt{F})$.
- Следовательно, некоторая достаточно длинная строка из нечетного количества букв $\mathbf{A}$ сопоставится с образцом P_i и войдет в $\mathscr{L}(\mathtt{F})$.

Образцы произвольного вида

- Если P допускает повторные вхождения переменных, $\mathscr{L}(P)$ есть сужение языка образцов (Angluin, 1980), описываемых P, на множество строк.
- $\mathscr{L}(P)$ не обязательно регулярны и даже контекстно-свободны. Пример: $P = e.1\,e.1$. Язык таких образцов является подмножеством индексных языков (indexed languages).
- Есть регулярные языки, не распознаваемые ни таким образцом, ни даже конечным их набором (Kari, 1995). Например, $\mathbf{A}^*\,\mathbf{B}$.
- Язык образцов не замкнут относительно ни одной из операций: $\cup$, $\cap$, $\setminus$.
- Проблема сводимости образцов произвольного вида неразрешима (Jiang, 1995).

Образцы произвольного вида

- Запрещаем подстановки пустой строки вместо переменной $\Rightarrow$ переход к *образцам с нестирающими подстановками*. Задача эквивалентности образцов в этом случае решается проверкой, существует ли переименование переменных (Angluin, 1980).
- Для нестирающих подстановок сводимость двух образцов одинаковой длины и сводимость произвольного образца к образцу, содержащему только одну переменную (произвольной кратности) также эквивалентна существованию сводящей подстановки.
- Если стирающие подстановки разрешены — *разрешимость задачи эквивалентности образцов под вопросом*!

Языки, распознаваемые функциями с произвольными образцами

Вопрос: Имея $\mathtt{F}_1$ и $\mathtt{F}_2$, можно ли построить функции, распознающие языки $\mathscr{L}(\mathtt{F}_1) \cup \mathscr{L}(\mathtt{F}_2)$, $\Sigma^* \setminus \mathscr{L}(\mathtt{F}_1)$?

Языки — пересечения языков образцов

$$
\begin{array}{l}
\mathtt{F}_1\,\{ \\
\qquad \mathsf{e.1}\ \mathsf{e.1}\ \mathbf{B} \quad = \mathbf{T}; \\
\qquad \mathsf{e.}z \qquad\qquad = \mathbf{F}; \quad \}
\end{array}
$$

$$
\begin{array}{l}
\mathtt{F}_2\,\{ \\
\qquad \mathsf{e.1}\ \mathbf{A}\ \mathbf{A}\ \mathsf{e.1}\ \mathbf{B} \quad = \mathbf{T}; \\
\qquad \mathsf{e.}z \qquad\qquad\qquad = \mathbf{F}; \quad \}
\end{array}
$$

Обе ветви **F** ограничений не порождают.

$\mathscr{L}(\mathtt{F}_1) \cap \mathscr{L}(\mathtt{F}_2) = (\mathbf{A}\,\mathbf{A})^{+}\,\mathbf{B}$, что можно понять, построив уравнение $\mathsf{e.1}\ \mathsf{e.1}\ \mathbf{B} = \mathsf{e.2}\ \mathbf{A}\ \mathbf{A}\ \mathsf{e.2}\ \mathbf{B}$. Оно сводится к уравнению $\mathbf{A}\ \mathsf{e.2} = \mathsf{e.2}\ \mathbf{A}$. Корни — $\mathsf{e.2} \in \mathbf{A}^{*}$.

Языки — пересечения языков образцов

Пусть P_1, P_2 — два произвольных образца (возможно, с пересекающимся множеством переменных).

Построим образец $\mathrm{P}_{AND} = \mathrm{P}_1\,\mathbf{A}\,\mathrm{P}_2\,\mathrm{P}_1\,\mathbf{B}\,\mathrm{P}_2$ и образец $\mathrm{P}_{TEST} = \mathrm{e.1}\,\mathbf{A}\,\mathrm{e.1}\,\mathrm{e.1}\,\mathbf{B}\,\mathrm{e.1}$.

$\mathscr{L}(\mathrm{P}_{AND}) \cap \mathscr{L}(\mathrm{P}_{TEST}) = \{\mathrm{e.1}\,\mathbf{A}\,\mathrm{e.1}\,\mathrm{e.1}\,\mathbf{B}\,\mathrm{e.1} : \mathrm{e}_1 = \mathrm{P}_1 \;\&\; \underline{\mathrm{P}_1 = \mathrm{P}_2}\}$.

$\mathrm{P}_1 = \mathrm{P}_2$ есть произвольное уравнение в словах.

Языки, распознаваемые образцами, содержащими запятую

При пересечении языков образцов порождаются *языки решений уравнений в словах*.

Уравнение в словах — равенство $\mathsf{P}_1 = \mathsf{P}_2$. Язык решений уравнения в словах $\mathscr{L}_{\mathsf{e}.x}(\mathsf{P}_1 = \mathsf{P}_2)$ по переменной $\mathsf{e}.x$ — множество строк, при подстановке которых в $\mathsf{P}_1 = \mathsf{P}_2$ вместо $\mathsf{e}.x$ оно не превращается в противоречивое.

Пример: Дано уравнение E: $\mathbf{A}\,\mathsf{e}.x\,\mathsf{e}.y = \mathsf{e}.x\,\mathsf{e}.y\,\mathbf{A}$. Рассмотрим язык его решений по $\mathsf{e}.x$. Если подставить вместо $\mathsf{e}.x$ в E любую строку, за исключением строк в языке $\mathbf{A}^*$, получится уравнение с пустым множеством решений. При подстановке вместо $\mathsf{e}.x$ строки из $\mathbf{A}^*$ получается уравнение на переменную $\mathsf{e}.y$, имеющее решение. Поэтому $\mathscr{L}_{\mathsf{e}.x}(E) = \mathbf{A}^*$.

Языки, распознаваемые функциями, содержащими запятую в образцах

- Замкнуты относительно операций $\setminus$, $\cup$, $\cap$.
- Включают как подкласс языки решений уравнений в словах.

Язык $\Sigma^* \setminus \mathscr{L}(\text{e.1}\,\mathbf{A}\,\text{e.2})$ не является языком решений никакого уравнения, но выразим с помощью набора Рефал-образцов.

Языки, распознаваемые функциями, содержащими запятую в образцах

$$
\texttt{F}\,\{
\begin{array}{ll}
\mathrm{e}.1,\ \mathrm{e}.1\,\mathrm{e}.1:\ \mathrm{s}.1\,\mathrm{e}.2\,\mathrm{e}.1\,\mathrm{e}.3,\ \mathrm{e}.1:\ \mathrm{s}.1\,\mathrm{e}.2\,\mathrm{s}.2\,\mathrm{e}.01 & = \mathbf{T}; \\
\mathrm{e}.z & = \mathbf{F}; \quad \}
\end{array}
$$

Функция F распознает т.н. *не простые* слова — такие значения u, что

$$\exists q, n(q \in \Sigma^+ \mathbin{\&} n \in \mathbb{N} \mathbin{\&} u = q^{n+2}).$$

Открытые вопросы

- Какие классы языков порождаются разными группами Рефал-образцов при неограниченном использовании в них $\mathbb{s}$-переменных?
- В каких частных случаях проблемы эквивалентности и вложения языков образцов Рефала разрешимы?