Уравнения с регулярными коэффициентами

Предварительные сведения.

Пусть p, q и r — регулярные выражения. Тогда справедливы следующие соотношения (употребляемые ниже знаки $|$ и $+$ равнозначны):

(1)	$p\|q = q\|p$;	(7)	$pe = ep = p$;
(2)	$\varnothing^* = e$;	(8)	$\varnothing p = p\varnothing = \varnothing$;
(3)	$p\|(q\|r) = (p\|q)\|r$;	(9)	$p^* = p\|p^*$;
(4)	$p(qr) = (pq)r$;	(10)	$(p^*)^* = p^*$;
(5)	$p(q\|r) = pq\|pr$;	(11)	$p\|p = p$;
(6)	$(p\|q)r = pr\|qr$;	(12)	$p\|\varnothing = p$.

* * *

При работе с языками часто бывает удобно пользоваться уравнениями, коэффициентами и неизвестными которых служат множества. Мы рассмотрим здесь системы уравнений, коэффициенты которых – регулярные выражения. Такие уравнения будем называть *уравнениями с регулярными коэффициентами*.

Рассмотрим, например, уравнение с регулярными коэффициентами

$$X = aX + b \tag{2.2.1}$$

где a и b – регулярные выражения. Легко проверить прямой подстановкой, что $X = a^*b$ – решение уравнения (2.2.1). Иначе говоря, если в обе части уравнения (2.2.1) подставить a^*b вместо X, то слева и справа будет одно и то же множество.

Можно рассматривать также системы уравнений, определяющие языки. Возьмём, например, пару уравнений

$$X = a_1X + a_2Y + a_3 \tag{2.2.2}$$

$$Y = b_1X + b_2Y + b_3$$

где a_i и b_i для всех $i = 1, 2, 3$ являются регулярными выражениями. Покажем, как решить эту систему уравнений и получить решение

$$X = (a_1 + a_2b_2^*b_1)^*(a_3 + a_2b_2^*b_3)$$

$$Y = (b_2 + b_1a_1^*a_2)^*(b_3 + b_1a_1^*a_3)$$

Однако сначала отметим, что не все уравнения с регулярными коэффициентами обладают единственным решением. Например, если

$$X = \alpha X + \beta \tag{2.2.3}$$

– уравнение с регулярными коэффициентами и α обозначает множество, содержащее пустую цепочку, то $X = \alpha^*(\beta + \gamma)$ будет решением уравнения (2.2.3) для любого γ (γ не обязано даже быть регулярным, см. упр. 2.2.7). Таким образом, уравнение (2.2.3) имеет бесконечно много решений. В такого рода ситуациях мы будем брать наименьшее решение, которое назовём *наименьшей неподвижной точкой*. Наименьшая неподвижная точка уравнения (2.2.3) – это множество $X = \alpha^*\beta$.

Определение. Систему уравнений с регулярными коэффициентами назовём *стандартной* системой с множеством неизвестных $\Delta = \{X_1, X_2, \ldots, X_n\}$, если она имеет вид

$X_1 = \alpha_{10} + \alpha_{11}X_1 + \alpha_{12}X_2 + \ldots + \alpha_{1n}X_n$

$\ldots$

$X_n = \alpha_{n0} + \alpha_{n1}X_1 + \alpha_{n2}X_2 + \ldots + \alpha_{nn}X_n$ – где все α_{ij} – регулярные выражения в алфавите, не пересекающемся с Δ.

Коэффициентами уравнений являются выражения α_{ij}. Заметим, что если $\alpha_{ij} = \varnothing$ (такое регулярное выражение возможно), то в уравнении для X_i нет члена, содержащего X_j. Аналогично, если $\alpha_{ij} = \varepsilon$, то в уравнении для X_i член, содержащий X_j, – это просто X_j. Иными словами, $\varnothing$ играет роль коэффициента 0, а ε – роль коэффициента 1 в обычных линейных уравнениях.

Алгоритм 2.1. Решение стандартной системы уравнений с регулярными коэффициентами.

Вход. Стандартная система Q уравнений с регулярными коэффициентами в алфавите Σ и с множеством неизвестных $\Delta = \{X_1, X_2, \ldots, X_n\}$.

Выход. Решение системы Q в виде $X_1 = \alpha_1, \ldots, X_n = \alpha_n$, где α_i – регулярное выражение в алфавите Σ.

Метод. Метод напоминает решение системы линейных уравнений методом исключения Гаусса.

Шаг 1. Положить $i = 1$.

Шаг 2. Если $i = n$, перейти к шагу 4. В противном случае с помощью тождеств леммы 2.1 записать уравнение для X_i в виде $X_i = \alpha X_i + \beta$, где α – регулярное выражение в алфавите Σ, а β - регулярное выражение вида $\beta_0 + \beta_i X_{i+1} + \ldots + \beta_n X_n$, причём все β_i – регулярные выражения в алфавите Σ. (Мы увидим, что это всегда возможно.) Затем в правых частях уравнений для $X_{i+1}, \ldots, X_n$ заменить X_i регулярным выражением $\alpha^*\beta$.

Шаг 3. Увеличить i на 1 и вернуться к шагу 2.

Шаг 4. Записать уравнение для X_n в виде $X_n = \alpha X_n + \beta$, где α и β – регулярные выражения в алфавите Σ. (После выполнения шага 2 для каждого $i < n$ в правой части уравнения для X_i не будет неизвестных $X_1, \ldots, X_{i-1}$. В частности, на шаге 4 этим свойством будет обладать и уравнение для X_n). Перейти к шагу 5 (при этом $i = n$).

Шаг 5. Уравнение для X_i имеет вид $X_i = \alpha X_i + \beta$, где α и β – регулярные выражения в алфавите Σ. Записать на выходе $X_i = \alpha^*\beta$ и в уравнения для $X_{i-1}, \ldots, X_1$ подставить $\alpha^*\beta$ вместо X_i.

Шаг 6. Если $i = 1$, остановиться. В противном случае уменьшить i на 1 и вернуться к шагу 5.

Пример 2.9. Пусть $\Delta = \{X_1, X_2, X_3\}$. Рассмотрим систему уравнений

$$X_1 = 0X_2 + 1X_1 + \varepsilon \tag{2.2.4}$$

$$X_2 = 0X_3 + 1X_2 \tag{2.2.5}$$

$$X_3 = 0X_1 + 1X_3 \tag{2.2.6}$$

Из уравнения (2.2.4) получаем $X_1 = 1X_1 + (0X_2 + \varepsilon)$. Затем в остальные уравнения подставляем $1^*(0X_2 + \varepsilon)$ вместо X_1. Уравнение (2.2.6) принимает вид

$$X_3 = 01^*(0X_2 + \varepsilon) + 1X_3$$

или с учётом леммы 2.1

$$X_3 = 01^*0X_2 + 1X_3 + 01^* \tag{2.2.7}$$

Если теперь из уравнения (2.2.5), которое мы ещё не трогали, выразить X_2 через X_3 и в (2.2.7) подставить 1^*0X_3 вместо X_2, то получим

$$X_3 = (01^*01^*0 + 1)X_3 + 01^* \tag{2.2.8}$$

Теперь в алгоритме 2.1 мы дошли до шага 5. Из уравнения (2.2.8) находим решение для X_3:

$$X_3 = (01^*01^*0 + 1)^*01^* \tag{2.2.9}$$

Подстановка (2.2.9) в (2.2.5) даёт

$$X_2 = 0(01^*01^*0 + 1)^*01^* + 1X_2 \tag{2.2.10}$$

Так как X_3 не входит в уравнение (2.2.4), то оно не меняется. Затем решаем (2.2.10) и получаем

$$X_2 = 1^*0(01^*01^*0 + 1)^*01^* \tag{2.2.11}$$

Подставляем (2.2.11) в (2.2.4):

$$X_1 = 01^*0(01^*01^*0 + 1)^*01^* + 1X_1 + \varepsilon \tag{2.2.12}$$

Решением уравнения (2.2.12) будет

$$X_1 = 1^*(01^*0(01^*01^*0 + 1)^*01^* + \varepsilon) \tag{2.2.13}$$

Выход алгоритма 2.1 – равенства (2.2.9), (2.2.11) и (2.2.13).

Мы должны показать, что выход алгоритма 2.1 действительно является решением данной системы уравнений в том смысле, что если подставить решение вместо неизвестных, то в каждом уравнении обе его части будут обозначать одно и то же множество. Как уже указывалось, решение стандартной системы уравнений не всегда единственно. Однако мы увидим, что в таком случае алгоритм 2.1 даёт наименьшую неподвижную точку.

Определение. Пусть Q – стандартная система уравнений с множеством неизвестных Δ и с коэффициентами в алфавите Σ. Отображение f множества Δ в множество языков в алфавите Σ называют *решением* системы Q, если после подстановки в каждое уравнение $f(X)$ вместо X для каждого $X \in \Delta$ уравнения становятся равенствами множеств. Отображение $f : \Delta \to P(\Sigma^*)$ называют *наименьшей неподвижной точкой* системы Q, если f – решение и для любого другого решения g

$f(X) \subseteq g(X)$ для всех $X \in \Delta$

Следующие две леммы содержат полезные сведения о наименьших неподвижных точках.

Лемма 2.2. *Каждая стандартная система уравнений Q с неизвестными Δ обладает единственной наименьшей неподвижной точкой.*

Доказательство. Пусть $f(X) = \{w | w \in g(X)$ для всех решений g системы $Q\}$ для всех $X \in \Delta$. Можно прямо показать, что f – решение и $f(X) \subseteq g(X)$ для всех решений g. Таким образом, f – единственная наименьшая неподвижная точка системы Q.

Теперь дадим некоторую характеристику наименьшей неподвижной точки.

Лемма 2.3. *Пусть Q – стандартная система уравнений с неизвестными $\Delta = X_1 \ldots X_n$ и уравнение для X_i имеет вид*

$X_i = \alpha_{i0} + \alpha_{i1}X_1 + \alpha_{i2}X_2 + \ldots + \alpha_{in}X_n$

Тогда наименьшей неподвижной точкой системы Q будет такое отображение f, что $f(X_i) = \{w_1 \dots w_m | w_m \in \alpha_{i_m 0}$ и $w_k \in \alpha_{j_k j_{k+1}}$ для некоторой последовательности чисел $j_1, \dots, j_m$, где $m \geqslant 1, 1 \leqslant k \leqslant m$ и $j_1 = i\}$.

Доказательство. Легко проверить непосредственно, что для всех i

$$f(X_i) = \alpha_{i0} \cup \alpha_{i1} f(X_1) \cup \dots \cup \alpha_{in} f(X_n)$$

Таким образом, f – решение.

Чтобы показать, что f – наименьшая неподвижная точка, предположим, что g – решение и для некоторого i существует цепочка $w \in f(X_i) - g(X_i)$. Так как $w \in f(X_i)$, то можно записать $w = w_1 \dots w_m$, где для некоторой последовательности чисел $j_1, \dots, j_m$ выполнены условия $w_m \in \alpha_{i_m 0}$ и $w_k \in \alpha_{j_k j_{k+1}}$ для $1 \leqslant k \leqslant m$ и $j_1 = i$. Так как g – решение, то

$g(X_j) = \alpha_{j0} \cup \alpha_{j1}\, g(X_1) \cup \dots \cup \alpha_{jn}\, g(X_n)$ для всех j

В частности, $\alpha_{j0} \subseteq g(X_j)$ и $\alpha_{jk}\, g(X_k) \subseteq g(X_j)$ для всех j и k. Тогда $w_m \in g(X_{j_m}), w_{m-1} w_m \in g(X_{j_{m-1}})$ и т.д. В конечном счёте получаем, что цепочки $w = w_1 w_2 \dots w_m$ принадлежат множеству $g(X_{j_1}) = g(X_i)$. Пришли к противоречию, так как предположили, что w не принадлежит $g(X_i)$. Таким образом, заключаем, что $f(X_i) \subseteq g(X_i)$ для всех i.

Отсюда непосредственно следует, что f – наименьшая неподвижная точка системы Q.

Лемма 2.4. *Пусть Q_1 и Q_2 – системы уравнений до и после одного применения шага 2 алгоритма 2.1. Тогда Q_1 и Q_2 имеют одну и ту же наименьшую неподвижную точку.*

Доказательство ...

Лемма 2.5. *Пусть Q_1 и Q_2 – системы уравнений до и после одного применения шага 5 алгоритма 2.1. Тогда Q_1 и Q_2 имеют одну и ту же наименьшую неподвижную точку.*

Доказательство (предоставлялось читателям).

Теорема 2.1. *Алгоритм 2.1 находит наименьшую неподвижную точку стандартной системы уравнений.*

Доказательство. После того как шаг 5 применён для всех i, все уравнения принимают вид $X_i = \alpha_i$, где α_i – регулярное выражение в алфавите Σ. Ясно, что отображение f, для которого $f(X_i) = \alpha_i$, и будет наименьшей неподвижной точкой этой системы.

ЛИТЕРАТУРА

1. *Ахо А., Ульман Дж.* Теория синтаксического анализа, перевода и компиляции. М.:Мир, 1978 г. Т. 1. С. 126–131.