Строене на регулярен израз по тотален КДА

Иво Стратев

9 юни 2020 г.

1 Конгругентност

Нека Σ е крайна азбука. Въвеждаме релация между регулярните изрази над Σ.

$$\mathsf{r} \approx_\Sigma \mathsf{s} \iff \mathcal{RL}_\Sigma(\mathsf{r}) = \mathcal{RL}_\Sigma(\mathsf{s})$$

В сила са

$$(\forall \mathsf{r}, \mathsf{s} \in \mathsf{RegExp}(\Sigma))(\mathsf{r} \approx_\Sigma \mathsf{s} \implies \mathsf{r}^* \approx_\Sigma \mathsf{s}^*)$$

$$(\forall \mathsf{r}, \mathsf{s}, \mathsf{t}, \mathsf{p} \in \mathsf{RegExp}(\Sigma))(\mathsf{r} \approx_\Sigma \mathsf{s} \;\&\; \mathsf{t} \approx_\Sigma \mathsf{p} \implies \mathsf{r}.\mathsf{t} \approx_\Sigma \mathsf{s}.\mathsf{p})$$

$$(\forall \mathsf{r}, \mathsf{s}, \mathsf{t}, \mathsf{p} \in \mathsf{RegExp}(\Sigma))(\mathsf{r} \approx_\Sigma \mathsf{s} \;\&\; \mathsf{t} \approx_\Sigma \mathsf{p} \implies \mathsf{r} + \mathsf{t} \approx_\Sigma \mathsf{s} + \mathsf{p})$$

$$(\forall \mathsf{r}, \mathsf{s} \in \mathsf{RegExp}(\Sigma))(\mathsf{r} \approx_\Sigma \mathsf{s} \implies \mathsf{r} + \mathsf{s} \approx_\Sigma \mathsf{s} + \mathsf{r})$$

$$(\forall \mathsf{r}, \mathsf{s}, \mathsf{t} \in \mathsf{RegExp}(\Sigma))((\mathsf{r} + \mathsf{s}).\mathsf{t} \approx_\Sigma (\mathsf{r}.\mathsf{t}) + (\mathsf{s}.\mathsf{t}))$$

$$(\forall \mathsf{r}, \mathsf{s}, \mathsf{t} \in \mathsf{RegExp}(\Sigma))(\mathsf{t}.(\mathsf{r} + \mathsf{s}) \approx_\Sigma (\mathsf{t}.\mathsf{r}) + (\mathsf{t}.\mathsf{s}))$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))(\emptyset + r \approx_\Sigma r)$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))(\emptyset.r \approx_\Sigma \emptyset)$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))(r.\emptyset \approx_\Sigma \emptyset)$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))(\varepsilon.r \approx_\Sigma r)$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))(r.\varepsilon \approx_\Sigma r)$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))((r^*)^* \approx_\Sigma r)$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))((\varepsilon + r)^* \approx_\Sigma r)$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))(r^*.(\varepsilon + r) \approx_\Sigma r^*)$$

$$(\forall r \in \mathsf{RegExp}(\Sigma))((\varepsilon + r).r^* \approx_\Sigma r^*)$$

2 Теорема на Клини

Теоремата на Клини гласи, че за всеки автоматен език съществува регулярен израз, който описва езика. Тоест класът на автоматните езици се включва в класът на регулярните езици.

Доказателството на теоремата дава алгоритъм за намиране на регулярен израз описващ език на даден креан тотален детерминиран автомат.

Нека $n \in \mathbb{N}_+$ и нека $Q = \{0, 1, \ldots, n-1\}$. Нека $\mathcal{A} = \langle \Sigma, Q, s, \delta, F \rangle$ е КТДА. Тогава можем да намерим регулярен израз r^k_{ij} съответстващ на езика R^k_{ij}, където $i, j \in Q$ и $k \in \{0, 1, \ldots n\}$, който описва думите, за които изичислението по автомата започнато от състояние i завършва в състояние j и преминава през междинни състояния само от множеството $\{0, 1, \ldots, k-1\}$.

Като

$$R^0_{ii} = \{\varepsilon\} \cup \{u \in \Sigma \mid i = \delta(i, u)\}$$

$$i \neq j \implies R^0_{ij} = \{u \in \Sigma \mid j = \delta(i,u)\}$$
$$R^{k+1}_{ij} = R^k_{ij} \cup R^k_{ik} \cdot (R^k_{kk})^* \cdot R^k_{kj}$$

Така $\mathcal{L}(\mathcal{A}) = \bigcup\limits_{j\in F} R^n_{0j}$.

3 Пример

Нека намерим регулярен израз описващ езика на автомата.

$$\begin{tikzcd}
\arrow[r] & 0 \arrow[loop above, "a"] \arrow[r, "b"] & 1 \arrow[loop above, "b"] \arrow[r, "a"] & 2 \arrow[loop above, "a,b"]
\end{tikzcd}$$

Правим следната таблица, съотвестваща на базовите случаи.

$k=0$	0	1	2
0	$\varepsilon + a$	b	$\emptyset$
1	$\emptyset$	$\varepsilon + b$	a
2	$\emptyset$	$\emptyset$	$\varepsilon + a + b$

След като използваме рекуретната връзка

$$R^{k+1}_{ij} = R^k_{ij} \cup R^k_{ik} \cdot (R^k_{kk})^* \cdot R^k_{kj}$$

Получаваме следната връзка $r^1_{ij} \approx r^0_{ij} + r^0_{i0} \cdot (r^0_{00})^* \cdot r^0_{0j}$, където $(r^0_{00})^* \approx (\varepsilon + a)^* \approx a^*$. Или така получаваме връзката $r^1_{ij} \approx r^0_{ij} + r^0_{i0} \cdot a^* \cdot r^0_{0j}$. Пресмятаме за $k = 1$.

$$r^1_{00} \approx r^0_{00} + r^0_{00} \cdot a^* \cdot r^0_{00} \approx (\varepsilon+a)+(\varepsilon+a)\cdot a^* \cdot (\varepsilon+a) \approx \varepsilon+a+(\varepsilon+a)\cdot a^* \approx \varepsilon+a+a^* \approx a^*$$

$$r^1_{01} \approx r^0_{01} + r^0_{00} \cdot a^* \cdot r^0_{01} \approx b+(\varepsilon+a)\cdot a^* \cdot b \approx b+a^* \cdot b \approx \varepsilon.b+a^* \cdot b \approx (\varepsilon+a^*)\cdot b \approx a^* \cdot b$$

$$r^1_{02} \approx r^0_{02} + r^0_{00} \cdot a^* \cdot r^0_{02} \approx \emptyset + (\varepsilon + a) \cdot a^* \cdot \emptyset \approx \emptyset + \emptyset \approx \emptyset$$

$$r^1_{10} \approx r^0_{10} + r^0_{10} \cdot a^* \cdot r^0_{00} \approx \emptyset + \emptyset \cdot a^* \cdot (\varepsilon + a) \approx \emptyset + \emptyset \approx \emptyset$$

$$r^1_{11} \approx r^0_{11} + r^0_{10} \cdot a^* \cdot r^0_{01} \approx \varepsilon + b + \emptyset \cdot a^* \cdot b \approx \varepsilon + b + \emptyset \approx \varepsilon + b$$

$$r^1_{12} \approx r^0_{12} + r^0_{10} \cdot a^* \cdot r^0_{02} \approx a + \emptyset \cdot a^* \cdot \emptyset \approx a + \emptyset \approx a$$
$$r^1_{20} \approx r^0_{20} + r^0_{20} \cdot a^* \cdot r^0_{00} \approx \emptyset + \emptyset \cdot a^* \cdot (\varepsilon + a) \approx \emptyset + \emptyset \approx \emptyset$$
$$r^1_{21} \approx r^0_{21} + r^0_{20} \cdot a^* \cdot r^0_{01} \approx \emptyset + \emptyset \cdot a^* \cdot b \approx \emptyset + \emptyset \approx \emptyset$$
$$r^1_{22} \approx r^0_{22} + r^0_{20} \cdot a^* \cdot r^0_{02} \approx \varepsilon + a + b + \emptyset \cdot a^* \cdot \emptyset \approx \varepsilon + a + b$$

Така получаваме таблицата

$k = 1$	0	1	2
0	a^*	$a^*.b$	$\emptyset$
1	$\emptyset$	$\varepsilon + b$	a
2	$\emptyset$	$\emptyset$	$\varepsilon + a + b$

От рекуретната връзка получаваме връзката

$$r^2_{ij} \approx r^{1+1}_{ij} \approx r^1_{ij} + r^1_{i1} \cdot (r^1_{11})^* \cdot r^1_{1j} \approx^1_{ij} + r^1_{i1} \cdot b^* \cdot r^1_{1j}$$

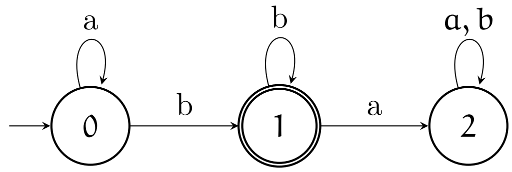
Пресмятаме за $k = 2$.

$$r^2_{00} \approx r^1_{00} + r^1_{01} \cdot b^* \cdot r^1_{10} \approx a^* + a^* \cdot b \cdot b^* \cdot \emptyset \approx a^*$$
$$r^2_{01} \approx r^1_{01} + r^1_{01} \cdot b^* \cdot r^1_{11} \approx a^* \cdot b^* + a^* \cdot b \cdot b^* \cdot (\varepsilon + b) \approx a^* \cdot b + a^* \cdot b \cdot b^* \approx a^* \cdot b \cdot (\varepsilon + b^*) \approx a^* \cdot b \cdot b^*$$
$$r^2_{02} \approx r^1_{02} + r^1_{01} \cdot b^* \cdot r^1_{12} \approx \emptyset + a^* \cdot b \cdot b^* \cdot a \approx a^* \cdot b \cdot b^* \cdot a$$
$$r^2_{10} \approx r^1_{10} + r^1_{11} \cdot b^* \cdot r^1_{10} \approx \emptyset + (\varepsilon + b) \cdot b^* \cdot \emptyset \approx \emptyset$$
$$r^2_{11} \approx r^1_{11} + r^1_{11} \cdot b^* \cdot r^1_{11} \approx (\varepsilon + b) + (\varepsilon + b) \cdot b^* + (\varepsilon + b) \approx (\varepsilon + b) + (\varepsilon + b) \cdot b^* \approx (\varepsilon + b) + b^* \approx b^*$$
$$r^2_{12} \approx r^1_{12} + r^1_{11} \cdot b^* \cdot r^1_{12} \approx a + (\varepsilon + b) \cdot b^* \cdot a \approx a + b^* \cdot a \approx (\varepsilon + b^*) \cdot a \approx b^* \cdot a$$
$$r^2_{20} \approx r^1_{20} + r^1_{21} \cdot b^* \cdot r^1_{10} \approx \emptyset + \emptyset \cdot b^* \cdot \emptyset \approx \emptyset$$
$$r^2_{21} \approx r^1_{21} + r^1_{21} \cdot b^* \cdot r^1_{11} \approx \emptyset + \emptyset \cdot b^* \cdot (\varepsilon + b) \approx \emptyset$$
$$r^2_{22} \approx r^1_{22} + r^1_{21} \cdot b^* \cdot r^1_{12} \approx (\varepsilon + a + b) + \emptyset \cdot b^* \cdot a \approx \varepsilon + a + b$$

Така получаваме таблицата

$k = 2$	0	1	2
0	a^*	$a^* \cdot b \cdot b^*$	$a^* \cdot b \cdot b^* \cdot a$
1	$\emptyset$	b^*	$b^* \cdot a$
2	$\emptyset$	$\emptyset$	$\varepsilon + a + b$

Понеже автоматът е



Той има едно финално състояние и значи $\mathcal{L}(\mathcal{A}) = R^3_{01} = \mathcal{RL}_{\{a,b\}}(r^3_{01})$. Следователно трябва да пресметнем само r^3_{01}.

$$r^3_{01} \approx r^2_{01} + r^1_{02} \cdot (r^2_{22})^* \cdot r^2_{21} \approx a^* \cdot b \cdot b^* + r^1_{02} \cdot (r^2_{22})^* \cdot \emptyset \approx a^* \cdot b \cdot b^* + \emptyset \approx a^* \cdot b \cdot b^*$$

Следователно $\mathcal{L}(\mathcal{A}) = \{a\}^* \cdot \{b\}^+$ и търсеният регулярен израз е $a^* \cdot b \cdot b^*$.