دانشکده‌ی علوم ریاضی

نظریه زبان‌ها و اتومات‌ها ۱۸ مهر ۱۳۹۱

جلسه‌ی ۸: تعیین عبارت منظم برای DFAها
قواعد جبری عبارت‌های منظم

مدرّس: دکتر شهرام خزائی *نگارندگان: نسرین راستگو، حمید احمدیان*

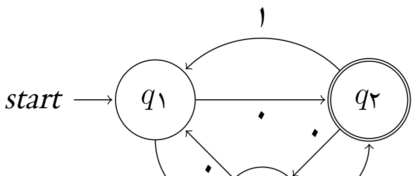
۱ تعیین عبارت منظم برای DFA

تعریف ۱ (مسیر) *در یک گراف انتقال حالت دنباله $(q_1, ..., q_{t+1})$ از حالت‌ها، مسیر متناظر با رشته $w = a_1 ... a_t$ نامیده می‌شود اگر $\delta(q_i, a_i) = q_{i+1}$.*

تعریف ۲ (k−مسیر) *فرض کنید A یک DFA با مجموعه حالت‌های $\{q_1, ..., q_n\}$ باشد. برای $k = 0, 1, ..., n$، یک k−مسیر برای این DFA، مسیری است که از هیچ حالت میانی با اندیس بزرگ‌تر از k نگذرد. (هیچ محدودیتی برای ابتدا و انتها وجود ندارد.)*

مثال ۱ *برای DFA مقابل k−مسیرها به شرح زیر هستند:*

- *مسیر (q_1) که متناظر با رشته‌ی ϵ است، یک ۰−مسیر، ۱−مسیر، ۲−مسیر و ۳−مسیر است.*
- *مسیر $(q_3 q_1 q_3)$ که متناظر با رشته‌ی 10 است، ۲−مسیر و ۱−مسیر است، اما ۰−مسیر نیست.*
- *همه‌ی مسیرها، ۳−مسیر هستند.*



تعریف ۳ *$R_{ij}^{(k)}$ عبارت منظمی است که زبانی را توصیف می‌کند که شامل همه‌ی رشته‌های متناظر با k−مسیرهایی است که از q_i شروع می‌شوند و به q_j ختم می‌شوند را تعیین می‌کند.*

دقت کنید که عبارت منظم $R_{ij}^{(k)}$ زیر زبانی را توصیف می‌کند است که در گراف انتقال حالت، q_i حالت آغازین است و q_j حالت نهایی.

مثال ۲ *می‌خواهیم عبارت منظمی برای مثال قبلی پیدا کنیم با با فرض اینکه $q_۳$ حالت نهایی باشد. در واقع ما به دنبال $R_{۱۳}^{(۳)}$ هستیم، یعنی تمام مسیرهایی که از $q_۱$ شروع می‌شوند و به $q_۳$ ختم می‌شوند. برای این کار روش بازگشتی زیر را بیان می کنیم:*
قبل از هر چیز چند تا از $R_{ij}^{(k)}$ ها را بدست می‌آوریم:

$$R_{23}^{(0)} = 0$$
$$R_{23}^{(1)} = 11 + 0$$
$$R_{23}^{(2)} = 1(01)^*1 + (10)^*0 = \{1(01)^n1 : n \geq 0\} \wedge \{(10)^n0 : n \geq 0\}$$

$R_{۲۳}^{(۳)}$ را نیز به روش بازگشتی می نویسم.

قضیه ۱ *برای هر DFA ایی یک عبارت منظم R وجود دارد که $L(R) = L(A)$.*
برهان. *فرض کنید مجموعه حالات $\{q_۱, ..., q_n\}$، حالت شروع $q_۱$ و مجموعه حالت‌های نهایی F باشد.*
از $R_{ij}^{(k)}$ جهت نام گذاری عبارت منظم استفاده می کنیم. برای ساخت عبارت $R_{ij}^{(k)}$ از روش استقرایی و با شروع از $k = ۰$ تا $k = n$ استفاده می‌کنیم. دقت کنید زمانی که $k = n$ هیچ محدودیتی روی مسیرها وجود ندارد ولی هیچ حالت با اندیس بزرگتر از n وجود ندارد.
به وضوح داریم:

$$R = \bigoplus_{q_j \in F} R_{۱j}^{(n)}$$

حال نشان می‌دهیم که $R_{ij}^{(k)}$ را چگونه می‌توان به صورت بازگشتی محاسبه کرد.
برای ساخت عبارت $R_{ij}^{(k)}$ از روش استقرایی استفاده می‌کنیم:
پایه: $k = ۰$
از آنجایی که اندیس تمام حالت‌ها ۱ یا بزرگتر هستند، محدودیت روی مسیرها این است که مسیر نباید هیچ حالت میانی داشته باشد. فقط دو نوع مسیر با این شرایط وجود دارد : مسیر به طول یک و مسیر به طول صفر.
بسته به مقادیر i و j دو حالت زیر را داریم:

۱. *$i \neq j$:*

در این حالت مسیر به طول صفر امکان ندارد و تنها مسیر به طول یک ممکن است وجود داشته باشد. اگر بین گره‌های q_i و q_j هیچ یالی وجود نداشته باشد، در این صورت مسیر به طول یک نیز وجود ندارد و بنابرین

$$R_{ij}^{(\circ)} = \emptyset$$

اگر بین q_i و q_j یالی با برچسب‌های $a_۱, ..., a_l \in \sum$ وجود داشته باشد، در این صورت l مسیر به طول یک داریم و بنابراین $R_{ij}^{(\circ)} = a_۱ + ... + a_l$. لذا در حالت کلی برای $i \neq j$ داریم:

$$R_{ij}^{(\circ)} = \bigoplus_{\delta(q_i,a)=q_j} a$$

۲. $i = j$:
در این حالت (q_i) خود یک مسیر به طول صفر (متناظر با رشته ϵ) است.
اگر گره q_i حلقه نداشته باشد، در این صورت $R_{ii}^{(0)} = \epsilon$.
اما اگر گره q_i دارای حلقه‌ای با برچسب‌های $a_1, ..., a_l$ باشد، در این صورت

$$R_{ii}^{(0)} = \epsilon + a_1 + ... + a_l$$

در حالت کلی داریم:

$$R_{ii}^{(0)} = \epsilon + \bigoplus_{\delta(q_j,a)=q_i} a$$

گام‌های استقرا:
فرض کنید یک مسیر از حالت q_i به حالت q_j وجود داشته باشد که در طول مسیر خود از هیچ حالت با اندیس بزرگتر از k نگذرد. دو امکان برای این حالت وجود دارد:

۱. مسیر از حالت q_k نگذرد: در این صورت زبان عبارت منظم $R_{ij}^{(ij)}$ تنها شامل همه رشته‌های متناظر با چنین مسیرهایی خواهد بود.

۲. مسیر حداقل یک بار از حالت q_k بگذرد: در این صورت ما می‌توانیم مسیر را به چند قسمت تقسیم کنیم. بخش اول از حالت i به حالت k است بدون گذر از k، قسمت آخر از k به حالت j است، و تمام حالت‌های میانی از k به خودش بدون گذر از k است. توجه کنید که اگر مسیر در طول خود فقط یک بار از k بگذرد، هیچ قسمت میانی‌ای وجود نخواهد داشت. عبارت منظم متناظر با تمام چنین مسیرهای به صورت زیر $R_{ik}^{(k-1)}(R_{kk}^{(k-1)})^*R_{kj}^{(k-1)}$ نشان داده می‌شود. عبارت اول بخشی از مسیر را نشان می‌دهد که تا ملاقات اولین k است، قسمت دوم مسیر با عبور از k به خودش را نشان می‌دهد، و قسمت سوم از k تا j را نشان می‌دهد.

از ترکیب عبارت‌های مسیرها در دو حالت فوق داریم:

$$R_{ii}^{(k)} = R_{ij}^{(k-1)} + R_{ik}^{(k-1)}(R_{kk}^{(k-1)})^*R_{kj}^{(k-1)}$$

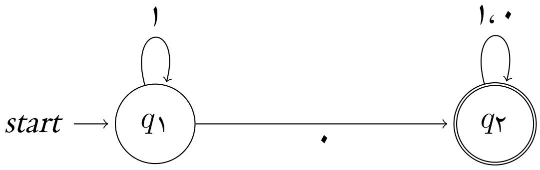
بنابراین $R_{ij}^{(n)}$ را به ازای هر i, j داریم. ■

۲ چند مثال

مثال ۳

$$R_{23}^{(3)} = R_{23}^{2} + R_{23}^{2}(R_{33}^{(2)})^*R_{33}^{2}$$

مثال ۴ برای DFA زیر یک عبارت منظم بنویسید.



این DFA تمام رشته‌هایی را که یک ۰ داشته باشند می‌پذیرد. زیرا، دقت کنید که این اتوماتا با شروع از q_1 و خواندن یک ۰ به حالت پایانی یعنی q_2 می‌رود و بعد به ازای هر رشته‌ی ورودی در q_2 باقی می‌ماند.

$$L = \{1^{(n)} 0 y : n \geq 0, y \in \{0, 1\}^*\}$$

حال می‌خواهیم عبارت منظم را براساس پایه‌های قضیه بسازیم.

Table 1: A DFA accepting all strings that have at least one 0

$R_{11}^{(0)}$	$\epsilon + 1$
$R_{12}^{(0)}$	0
$R_{21}^{(0)}$	$\emptyset$
$R_{22}^{(0)}$	$(\epsilon + 0 + 1)$

عبارت $R_{11}^{(0)}$ دارای $\emptyset$ است زیرا حالت ابتدایی و انتهایی یکسان است. دارای ۱ است زیرا یک یال از حالت q_1 به حالت q_1 با ورودی ۱ دارد. به عنوان مثال دیگر، $R_{12}^{(0)}$ ۰ است زیرا یک یال با ورودی ۰ از حالت q_1 به حالت q_2 وجود دارد. هیچ $\emptyset$ ندارد زیرا حالتهای ابتدایی و انتهایی متفاوت هستند. و به عنوان مثال سوم، $R_{21}^{(0)} = \emptyset$ است زیرا هیچ یالی از ۲ به ۱ وجود ندارد. حال می‌بایست پایه‌های استقرا را بسازیم. قاعده‌ی محاسبه‌ی $R_{ij}^{(0)}$ براساس قضیه برابر است با:

$$R_{ij}^{(1)} = R_{ij}^{(0)} + R_{i1}^{(0)}(R_{11}^{(0)})^* R_{1j}^{(0)}$$

در جدول زیر دومین ستون اولین عبارت‌های محاسبه شده مستقیم با استفاده از فرمول فوق است، و ستون سوم ساده شده‌ی عبارت‌ها است.

Table 2: Regular expression for paths that can go through only state 1

	By direct substitution	***Simplified***
$R_{11}^{(1)}$	$\epsilon + 1 + (\epsilon + 1)(\epsilon + 1)^*(\epsilon + 1)$	1^*
$R_{12}^{(1)}$	$0 + (\epsilon + 1)(\epsilon + 1)^*0$	1^*0
$R_{21}^{(1)}$	$\emptyset + \emptyset(\epsilon + 1)^*(\epsilon + 1)$	$\emptyset$
$R_{22}^{(1)}$	$\epsilon + 0 + 1 + \emptyset(\epsilon + 1)^*0$	$\epsilon + 0 + 1$

برای نمونه، عبارت $R_{۱۲}^{(۱)}$ را در نظر بگیرید. این عبارت برابر است با $R_{۱۲}^{(۰)} + R_{۱۱}^{(۰)}(R_{۱۱}^{(۰)})^*R_{۱۲}^{(۰)}$. برای فهم عبارت‌های ساده شده توجه کنید که قاعده‌ی اگر R هر عبارت منظمی باشد، خواهیم داشت $(\epsilon + R)^* = R^*$. نکته این است که هر دو طرف عبارت زبانی را توضیح را می‌دهند که حاوی هر ترکیبی از ۰ یا عبارت‌های بیشتر از $L(R)$ است. در موقعیت ما، داریم $(\epsilon + ۱)^* = ۱^*$، توجه کنید که هر دو عبارت بیان می‌کنند که عبارت حاوی هر تعداد ۱ است و $(\epsilon + ۱)۱^* = ۱^*$ که بازهم نشان‌دهنده هر تعداد ۱ است. بنابراین عبارت اصلی $R_{۱۲}^{(۱)}$ برابراست با $۰ + ۱^*۰$. ساده شده‌ی عبارت $R_{۱۱}^{(۱)}$ شبیه عبارت قبلی است. ساده شده‌ی عبارت $R_{۲۱}^{(۱)}$ و $R_{۲۲}^{(۱)}$ بستگی به دو قاعده در مورد $\emptyset$ دارد. برای هر عبارت منظم R داریم:

۱. $\emptyset R = R\emptyset = \emptyset$. بنابراین عبارت $\emptyset$ عضوپوچ ساز الحاق[۱] است.

۲. $\emptyset + R = R + \emptyset = R$. بنابراین عبارت $\emptyset$ عضو همانی اجتماع[۲] است.

در نتیجه یک عبارت مانند $\emptyset(\epsilon + ۱)^*(\epsilon + ۱)$ با $\emptyset$ جایگزین می‌شود. حال می‌خواهیم عبارت $R_{ij}^{(۲)}$ را محاسبه کنیم. $k = ۲$ قرار می‌دهیم:

$$R_{ij}^{(2)} = R_{ij}^{(1)} + R_{i2}^{(1)}(R_{22}^{(1)})^*R_{2j}^{(1)}$$

Table 3: Regular expressions for paths that can go through any state 0

	By direct substitution	***Simplified***
$R_{11}^{(2)}$	$1^* + 1^*0(\epsilon + 0 + 1)^*\emptyset$	1^*
$R_{12}^{(2)}$	$1^*0 + 1^*0(\epsilon + 0 + 1)^*(\epsilon + 0 + 1)$	$1^*0(0 + 1)^*$
$R_{21}^{(2)}$	$\emptyset + (\epsilon + 0 + 1)(\epsilon + 0 + 1)^*\emptyset$	$\emptyset$
$R_{22}^{(2)}$	$\emptyset + 0 + 1 + (\epsilon + 0 + 1)(\epsilon + 0 + 1)^*(\epsilon + 0 + 1)$	$(0 + 1)^*$

عبارت منظم نهایی برای این اتوماتا از اجتماع تمام عبارت‌هایی که از حالت اولیه شروع و دومین حالت، حالت نهایی باشد، ساخته می‌شود. در این مثال، با شروع از $q_۱$ و تنها حالت نهایی $q_۲$، ما فقط به عبارت $R_{۱۲}^{(۲)}$ نیاز داریم. این عبارت برابر است با $۱^*۰(۰ + ۱)^*$. زبان این عبارت شامل تمام رشته‌هایی است با ۰ یا هر تعداد ۱ شروع می‌شود و بعد هر تعداد ۰ و ۱ خواهد داشت.

[۱] annihilator for concatenation

[۲] identity for union

۳ قواعد جبری عبارت‌های منظم

قاعده‌ی جابه‌جایی[۳]

$$E + F = F + E$$
$$EF \neq FE$$

قاعده‌ی اشتراک‌پذیری[۴]

$$(E + F) + R = E + (F + R)$$
$$(EF)R = E(FR)$$

قاعده‌ی پخشی[۵]

$$L(M + N) = LM + LN$$
$$(M + N)L = ML + NL$$

عضو همانی جمع

$$\emptyset + R = R + \emptyset = R$$

عضو همانی الحاق

$$\epsilon R = R\epsilon = R$$

عضو پوچ ساز الحاق

$$\emptyset R = R\emptyset = \emptyset$$

قاعده‌ی همانی برای جمع: این قاعده بیان می‌کند که اجتماع هر تعداد از عبارت‌های یکسان را می‌توان با یکی از آن عبارت‌ها جایگزین کرد.

$$R + R = R$$

بسته شدن عبارت $\emptyset$ تنها حاوی رشته‌ی ϵ است.

$$\emptyset^* = \epsilon$$

چک کردن درستی عبارت زیر ساده است. تنها رشته‌ای که با هر تعداد الحاق ϵ برابر با خودش و ϵ است.

$$\epsilon^* = \epsilon$$

[۳] commutative law
[۴] associative law
[۵] distributive law