دانشکده‌ی علوم ریاضی

نظریه‌ی زبان‌ها و اتوماتا ۲۵ مهر ۱۳۹۱

جلسه‌ی ۱۰: ویژگی‌های زبان منظم، نامحدود بودن زبان، لم تزریق، برابری زبان‌ها

مدرّس: دکتر شهرام خزائی *نگارندگان: نوید علامتی، محمد مبین جاریحانی*

۱ ویژگی‌های زبان منظم

ویژگی‌های زبان منظم شامل دو دسته می‌شود:

- ویژگی‌های تصمیمی [۱]
- ویژگی‌های بستاری [۲]

تعریف ۱ *ویژگی تصمیمی زبان‌های منظم، شامل بررسی الگوریتم‌هایی است که توصیف صوری [۳] (برای مثال DFA) یک یا چند زبان منظم را به عنوان ورودی می‌گیرد و مشخص می‌کند که آیا زبان‌های مورد نظر، برخی خواص مشخص را دارند یا خیر.*

برای مثال، آیا زبان منظم L تهی است یا خیر. آیا دو زبان منظم L و M برابر هستند یا خیر.

تعریف ۲ *ویژگی بستاری مشخص می‌کند که آیا منظم بودن زبان‌ها، تحت یک سری عملیات (برای مثال اجتماع یا مکمل‌گیری) حفظ می‌شود یا خیر.*

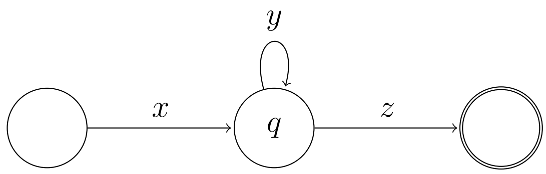
۲ مسأله‌ی نامحدود بودن زبان

آیا زبان منظم L نامحدود است؟
برای پاسخ به این سوال دو لم را بررسی می‌کنیم.

لم ۱ *اگر DFA زبان منظم L دارای n حالت باشد و L دارای رشته‌ای به طول n یا بیشتر باشد، آنگاه زبان L نامحدود است.*

[۱] decision properties
[۲] closure properties
[۳] formal description

برهان. اگر یک ماشین حالت متناهی معین به تعداد حالات n یک رشته‌ی فرضی w به طول n یا بیشتر را قبول کند، آنگاه بنا به اصل لانه‌ی کبوتری یک حالت فرضی q وجود دارد که دو بار در طول مسیر پردازش w ظاهر شده است. به عبارت دیگر، رشته‌ی w به صورت $w = xyz$ می‌باشد که $y \neq \epsilon$، به‌طوری که با پردازش پیشوند x از حالت آغازین به حالت q می‌رسیم و y رشته‌ای است که پس از پردازش آن با شروع از q دوباره به خود q رسیده‌ایم. پسوند z رشته‌ای است که ما را از حالت q به یک حالت نهایی برده است. واضح است که تمام رشته‌های xy^iz برای i هایی که $i \geq 0$ هستند، مورد قبول DFA می‌باشد. چون $y \neq \epsilon$ است، تمام این رشته‌ها متفاوت هستند و لذا تعداد آنها نامتناهی است.



■

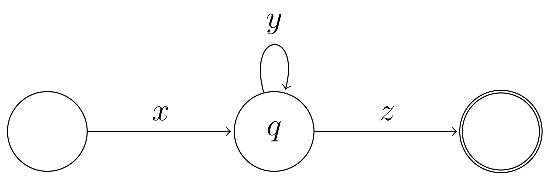
توجه شود که لم ۱ به تنهایی برای طراحی یک الگوریتم که متناهی بودن یا نبودن یک زبان را تعیین می‌کند کافی نیست. چون بی‌شمار رشته به طول بزرگتر از n داریم، اگر زبان داده شده متناهی باشد، الگوریتم فرضی هیچ‌گاه متوقف نخواهد شد. بنابراین به یک لم دیگر نیاز داریم که بتوان الگوریتمی بر مبنای آن طراحی کرد که تعداد متناهی رشته را امتحان کند.

لم ۲ *اگر زبان منظم L نامتناهی باشد و DFA آن دارای n حالت باشد، آنگاه L دارای رشته‌ای است که طولش بین n و $2n - 1$ می‌باشد.*

برهان. طبق فرض، رشته‌ی w به طول حداقل n در زبان L وجود دارد. اگر $|w| < 2n$ که حکم ثابت است. فرض کنید $|w| \geq 2n$، $w = xyz$ و y اولین دور در مسیرمان به سمت حالت نهایی باشد. پس داریم:

$$1 \leq |y| \leq n \quad , \quad |xy| \leq n$$

طبق لم ۱ رشته‌ی xz (در xy^iz قرار دهید $i = 0$) نیز در زبان L است. دقت کنید $n \leq |xz| < |w|$ اگر طول رشته‌ی جدید کمتر از $2n$ نباشد، به همین ترتیب می‌توان ادامه داد و به یک رشته که طولش بین n و $2n - 1$ است، رسید.



■

در نتیجه ما باید تمام رشته‌هایی را که طولشان بین n و $2n - 1$ باشد، پردازش کنیم. اگر به یکی از حالات نهایی رسیدیم زبان نامتناهی است (بنا به لم ۱) و اگر نه، متناهی نیست. ولی برای مشخص کردن نامتناهی بودن یک زبان باید تمام رشته‌هایی را که طولشان در بازه‌ی $[n, 2n - 1]$ است، پردازش کنیم که راه حل کارایی نیست.
یک الگوریتم کارا برای مسأله‌ی نامحدود بودن زبان ارائه می‌کنیم:

- حذف حالت‌هایی از DFA که از حالت آغازین قابل دستیابی نیستند.
- حذف حالت‌هایی از DFA که به یک حالت نهایی ختم نمی‌شوند.

- آیا گراف باقی‌مانده دارای دور است یا خیر.
- اگر دارای دور بود، زبان نامتناهی است و در غیر این صورت متناهی است.

۳ لم تزریق

در این بخش شرطی لازم(و نه لزوما کافی) برای منظم بودن یک زبان معرفی می‌کنیم. با این شرط می‌توان منظم نبودن بعضی زبان‌ها را اثبات کرد. ایده‌ی کار بسیار ساده است. کافی است زبان‌های نامتناهی را در نظر بگیریم زیرا هر زبان متناهی، منظم است. فرض کنیم زبان منظم نامتناهی L دارای DFA با n حالت باشد. بدیهی است که این زبان دارای کلماتی، مانند w، با طول بزرگ‌تر یا مساوی n است. چون این کلمه در زبان قرار دارد، دارای مسیر پذیرش از q_0 به یک حالت پایانی مانند q_f می‌باشد. طول این مسیر با طول w برابر است. طول بزرگ‌ترین مسیر در این DFA که خودش را قطع نمی‌کند(یعنی هیچ یک از حالات q_i دو بار در مسیر ظاهر نمی‌شود) حداکثر $n-1$ خواهد بود. بنابراین مسیر پذیرش w حداقل یک بار خودش را قطع می‌کند. فرض کنیم q_i اولین حالتی باشد که این مسیر خودش را قطع می‌کند. پس w را می‌توان به صورت xw' نوشت که در آن با مصرف قسمت x از q_0 به q_i می‌رسیم. طبق مطالب گفته شده، w' را می‌توان به صورت yz نوشت که در آن y با شروع از q_i، دوباره به آن بازمی‌گردد. بنابراین اگر y را حذف کنیم یا آن را تکرار کنیم(یعنی روی دور چندین بار حرکت کنیم) باز هم می‌توانیم به حالت پایانی برسیم. در واقع با تزریق(یا حذف) y کلمه‌ی حاصل باز هم در L قرار می‌گیرد. این موضوع در لم زیر بیان شده است.

لم ۳ *برای هر زبان منظم L عدد طبیعی n (که تعداد حالت های DFA متناظر با زبان است) وجود دارد که برای هر رشته‌ی $w \in L$ با طول بزرگتر یا مساوی n می‌توان رشته‌ی w را بصورت $w = xyz$ نوشت که سه شرط زیر برقرار باشند:*

- $|xy| \leq n$
- $|y| \geq 1$
- $\forall i \geq 0 : xy^iz \in L$

برای زبان منظم L با DFA دارای n حالت، این لم را می‌توان با استفاده از صورها به صورت زیر نوشت

$$\begin{aligned}
&\exists n \in \mathbb{N} \\
&\forall w \in L : |w| \geq n \\
&\exists x, y, z \in \{0, 1\}^* : \left\{ \begin{array}{c} w = xyz \\ |xy| \leq n \\ |y| \geq 1 \\ \forall i \geq 0 : xy^iz \in L \end{array} \right.
\end{aligned} \tag{1}$$

با گرفتن نقیض از عبارت بالا، شرطی کافی برای منظم نبودن یک زبان به دست می‌آید.

$$\begin{array}{l} \forall n \in \mathbb{N} \\ \exists w \in L : |w| \geq n \\ \forall x,y,z \in \{۰,۱\}^* : \neg \left\{ \begin{array}{c} w = xyz \\ |xy| \leq n \\ |y| \geq ۱ \\ \forall i \geq ۰ : xy^iz \in L \end{array} \right. \end{array} \tag{۲}$$

که معادلا می‌توان نوشت:

$$\begin{array}{l} \forall n \in \mathbb{N} \\ \exists w \in L : |w| \geq n \\ \forall x,y,z \in \{۰,۱\}^* : \neg \Big(\forall i \geq ۰ : xy^iz \in L \Big) \vee \neg \left\{ \begin{array}{l} w = xyz \\ |xy| \leq n \\ |y| \geq ۱ \end{array} \right. \end{array} \tag{۳}$$

که معادل است با:

$$\begin{array}{l} \forall n \in \mathbb{N} \\ \exists w \in L : |w| \geq n \\ \forall x,y,z \in \{۰,۱\}^* : \Big(\exists i \geq ۰ : xy^iz \notin L \Big) \vee \neg \left\{ \begin{array}{l} w = xyz \\ |xy| \leq n \\ |y| \geq ۱ \end{array} \right. \end{array} \tag{۴}$$

که نهایتا می‌توان آن را به صورت زیر نوشت (دقت کنید که $p \Rightarrow q$ معادل است با $q \vee \neg p$):

$$\begin{array}{l} \forall n \in \mathbb{N} \\ \exists w \in L : |w| \geq n \\ \forall x,y,z \in \{۰,۱\}^* : \left\{ \begin{array}{l} w = xyz \\ |xy| \leq n \\ |y| \geq ۱ \end{array} \right. \Rightarrow \Big(\exists i \geq ۰ : xy^iz \notin L \Big) \end{array} \tag{۵}$$

به بیان دیگر باید برای هر n داده شده کلمه‌ای با طول حداقل n مانند w پیدا کنیم که اگر آن را به هر طریق به xyz تجزیه کنیم، یک i وجود داشته باشد که نتوان y^i را به w تزریق کرد به طوری که کلمه‌ی xy^iz حاصل در L باشد. حال به بیان چند مثال می‌پردازیم.

مثال ۱ *زبان* $L = \{۰^k۱^k | k \geq ۰\}$ *منظم نیست.*

فرض کنیم L منظم باشد و DFA نظیر آن دارای n حالت باشد. قرار می‌دهیم $w = ۰^n۱^n$. هرگاه $w = xyz$ یک تجزیه با $|xy| \leq n$ و $y \neq \epsilon$ باشد، آنگاه xy و لذا y فقط شامل ۰ است. به وضوح با اضافه کردن(یا حذف) y کلمه‌ای با تعداد ۰ و ۱ نابرابر ایجاد می‌شود که در L نیست. این تناقض نشان می‌دهد که L نمی‌توان منظم باشد.

مثال ۲ *زبان $L = \{1^q | q \text{ is prime}\}$ منظم نیست.*

مانند مثال قبل فرض کنیم L منظم باشد و DFA نظیر آن دارای n حالت باشد. قرار می‌دهیم $w = 1^q$ که در آن q یک عدد اول با $q \geq n + 2$ است. حال طبق نقیض لم تزریق کافیست یک i بیابیم که xy^iz در L نباشد و این معادل آن است که طول این کلمه نباید عددی اول باشد. نشان می‌دهیم $i = q + 1$ در این شرط صدق می‌کند. در واقع $|xy^iz| = |x| + i|y| + |z| = |x| + |y| + |z| + (i - 1)|y| = q + q|y| = q(1 + |y|)$. به وضوح هر دو عامل حاصل‌ضرب ناسره هستند و لذا این عدد اول نیست. این منظم بودن L را نقض می‌کند. در این مثال انتخاب $q \geq n + 2$ برای آن است که هم شرط لم تزریق برقرار شود و هم $q > 1$ باشد که عدد حاصل اول نشود. با استفاده از این لم می‌توان نشان داد که زبان‌های $\{0^m 1^n; m > n\}$ و $\{ww; w \in \{0, 1\}^*\}$ نیز منظم نیستند.

۱.۳ بیان لم تزریق به صورت یک بازی

استفاده از لم تزریق را در یک بازی دونفره شرح می‌دهیم:
بازیکن ۱ یک زبان انتخاب می‌کند تا ثابت کند که زبان انتخاب شده منظم نیست.
بازیکن ۲ یک عدد n انتخاب می‌کند ولی n را فاش نمی‌کند، پس بازیکن ۱ باید به‌صورتی بازی کند که استدلالش برای تمامی n ها صحیح باشد.
بازیکن ۱ یک رشته‌ی دلخواه w را انتخاب می‌کند که طولش حداقل برابر n است.
بازیکن ۲، w را به سه زیر رشته‌ی xyz تقسیم می‌کند که در شرایط لم تزریق صدق می‌کند، یعنی:

$$y \neq \epsilon \quad , \quad |xy| \leq n$$

همچنین بازیکن ۲ نباید به بازیکن ۱ بگوید که xyz چیست اگرچه xyz باید در شرایط لم تزریق صدق کند.
اگر بازیکن ۱ یک عدد i معرفی کند که i ممکن است تابعی از x، y، z و n باشد، به‌طوری‌که xy^iz در L نباشد، آنگاه برنده است.

۴ مسأله‌ی برابری دو زبان

دو زبان L و M را داریم. برای پاسخ دادن به این سوال که آیا دو زبان L و M برابر هستند، از یک ابزار به نام DFA حاصل‌ضرب [۴] استفاده می کنیم.

تعریف ۳ *حاصل‌ضرب ماشین‌های حالت متناهی معین $A = (Q_A, \Sigma, \delta, q_A, F_A)$ و $B = (Q_B, \Sigma, \delta, q_B, F_B)$ به‌صورت $A \times B = (Q, \Sigma, \delta, q_0, F)$ تعریف می‌شود که:*

- *$Q = Q_A \times Q_B$ ، یعنی مجموعه‌ی تمام زوج‌های (q, r) که q متعلق به Q_A و r متعلق به Q_B است.*
- *تعریف تابع انتقال حالت:*

$$\forall a \in \Sigma, q \in Q_A, r \in Q_B : \delta((q, r), a) = (\delta_A(q, a), \delta_B(r, a))$$

- *حالت آغازین:* $q_0 = (q_A, q_B)$

[۴] product DFA

• *حالت نهایی: شامل تمام زوج‌هایی است که فقط یک مؤلفه‌ی آن‌ها در زبان خودش حالت نهایی باشد، یعنی:*

$$F = \{(q,r)|(q \in F_A \wedge r \notin F_B) \vee (q \notin F_A \wedge r \in F_B)\}$$

لم ۴ *دو زبان L و M برابرند، اگر و فقط اگر زبان DFA حاصل‌ضربشان تهی باشد.*

دلیل این شرط این است که ما اثبات می‌کنیم هیچ رشته‌ای وجود ندارد که در یکی باشد و در دیگری نباشد.
مثال : DFA زبان L، DFA زبان M و در نهایت DFA حاصل‌ضرب آن‌دو به ترتیب رسم شده‌اند. مشاهده می‌شود که زبان حاصل‌ضرب تهی نیست، بنابراین L و M برابر نیستند.