نظریه‌ی زبان‌ها و ماشین‌ها

درس‌نامه‌ی ۸

کاظم فولادی
http://kazim.fouladi.ir

ویراست اول: ۱۳۸۵
ویراست دوم: ۱۳۸۷
ویراست سوم: ۱۳۹۳

فهرست مطالب

۸ خصوصیات زبان‌های مستقل از متن

Properties of Context-Free Languages

۸-۱ لم تزریق و زبان‌های مستقل از متن

۸-۱-۱ لم تزریق برای زبان‌های مستقل از متن

قضیه

فرض کنید که L یک زبان مستقل از متن نامتناهی باشد. در این صورت عدد صحیح مثبت m وجود دارد که هر $w \in L$ با $|w| \geq m$ می‌تواند به صورت زیر تجزیه شود:

$$w = uvxyz$$

که در آن

$$|vxy| \leq m, \qquad |vy| \geq ۱$$

به طوری که به ازای هر $i = ۰, ۱, ۲, \ldots$، $uv^ixy^iz \in L$ است.

◄ **اثبات.**

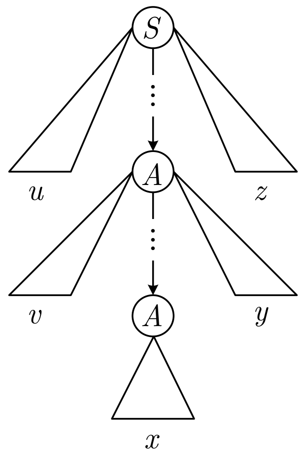
زبان $L - \{\lambda\}$ را در نظر می‌گیریم و فرض می‌کنیم گرامر G را برای آن داریم که عاری از قواعد λ و یکه است:

$$L(G) = L - \{\lambda\}$$

طول سمت راست هر قاعده محدود است، یعنی برای هر قاعده‌ی $A \to \alpha$ داریم $|\alpha| \leq k$،
پس طول هر اشتقاق از رشته‌ای مانند $w \in L$ باید حداقل $|w|/k$ باشد:

$$S \Rightarrow^n w, \qquad n \geq \frac{|w|}{k}$$

چون L نامتناهی است، اشتقاق‌هایی با طول دلخواه بلند وجود دارند،
پس درخت‌های اشتقاقی با ارتفاع دلخواه وجود خواهند داشت.
یک درخت اشتقاق با مسیری به طول کافی بلند از ریشه به یک برگ را در نظر می‌گیریم. چون تعداد متغیرهای G متناهی است، حتماً باید یک متغیر در این مسیر تکرار شود:



اشتقاق زیر معادل با درخت فوق را داریم:

$$S \Rightarrow^* uAz \Rightarrow^* uvAyz \Rightarrow^* uvxyz, \quad u,v,x,y,z \in T^*$$

از این اشتقاق به دست می‌آوریم:

$$A \Rightarrow^* vAy, \quad A \Rightarrow^* x$$

یعنی A به تعداد دلخواه می‌تواند اشتقاق $A \Rightarrow^* vAy$ را تکرار نماید، پس:

$$\forall i \in \{۰, ۱, ۲, \ldots\} \Rightarrow uv^ixy^iz \in L(G)$$

در اشتقاق‌های $A \Rightarrow^* vAy$ و $A \Rightarrow^* x$ می‌توان فرض کرد که هیچ متغیری تکرار نمی‌شود (در غیر این صورت متغیر تکراری را A در نظر می‌گیریم).
بنابراین، طول رشته‌های v، x و y فقط وابسته به قواعد گرامر است و مستقل از w قابل محدود شدن است، تا شرط $|vxy| \leq m$ برقرار شود.
چون قواعد یکه و λ نداریم، v و y نمی‌توانند هر دو λ باشند و بنابراین $|vy| \geq ۱$.

کاربرد لم تزریق، برای نشان دادن مستقل از متن نبودن یک زبان است. (مشابه لم تزریق زبان‌های منظم که برای اثبات منظم نبودن یک زبان به کار می‌رود.)

- در زبان‌های منظم، زیررشته‌ی xy با طول محدود به m از ابتدای w شروع می‌شود و رشته‌ی y که تزریق می‌شود، در m نماد اول w قرار دارد.
- اما در زبان‌های مستقل از متن، محدودیت فقط روی $|vxy|$ است و زیررشته‌ی u قبل از vxy می‌تواند به دلخواه طولانی باشد.
- در نتیجه، در بحث رقابتی، حریف آزادی عمل بیشتری خواهد داشت و بازی برای ما دشوارتر می‌شود. در واقع تعداد حالت‌های بیشتری برای تجزیه‌ی رشته قابل تصور است که همه‌ی آنها باید بررسی شوند.

مثال

نشان دهید که زبان $L = \{a^n b^n c^n : n \geq 0\}$ مستقل از متن نیست.

- **(حریف)** $m > 0$
- **(خودمان)** $|w| \geq m$ $\quad w = a^m b^m c^m \in L,$
- **(حریف)** تجزیه‌ی w با شرط $|vxy| \leq m, \ |vy| \geq 1$

۱) vxy فقط حاوی a باشد.

۲) vxy حاوی تعداد مساوی a و b انتخاب شود.

۳) vxy به گونه‌ای انتخاب شود که vy تعداد مساوی a، b و c داشته باشد (که ممکن نیست، زیرا باید $|vxy| \leq m$).

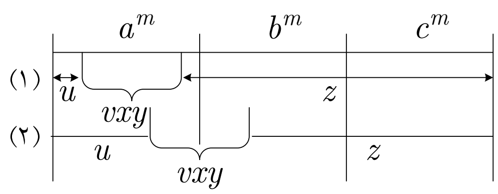
- **(خودمان)**

۱) رشته‌ی تزریق‌شده در L نیست.

۲) رشته‌ی تزریق‌شده $a^k b^k c^m$ با $k \neq m$ در L نیست.

۳) حریف در انتخاب تجزیه‌ی مناسب w برای شکست دادن ما موفق نمی‌شود.

پس زبان L یک زبان مستقل از متن نیست.

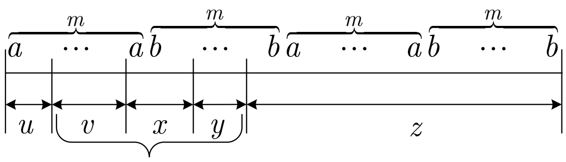


مثال

نشان دهید که زبان $L = \{ww : w \in \{a,b\}^*\}$ مستقل از متن نیست.

- **(حریف)** $m > 0$
- **(خودمان)** $|w| \geq m$ $\quad w = a^m b^m a^m b^m \in L,$
- **(حریف)** راه‌های مختلفی برای تجزیه‌ی رشته و انتخاب vxy وجود دارد.
  یکی از این راه‌ها مطابق شکل زیر است.
- **(خودمان)** در تجزیه‌ی شکل زیر داریم: $i = 0 \Rightarrow a^k b^j a^m b^m \notin L, \ k < m \text{ or } j < m$

انتخاب سایر تجزیه‌ها نیز به نتیجه‌ی مشابهی منجر می‌شود. بنابراین حریف در انتخاب تجزیه‌ی مناسب برای شکست دادن w ما موفق نمی‌شود؛
پس زبان L یک زبان مستقل از متن نیست.



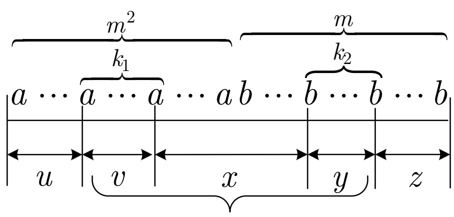
مثال

نشان دهید که زبان $L = \{a^n b^j : n = j^2\}$ مستقل از متن نیست.

- **(حریف)** $m > 0$
- **(خودمان)** $w = a^{m^2} b^m \in L, \quad |w| \geq m$
- **(حریف)** گزینه‌های مختلفی برای تجزیه و انتخاب xyz وجود دارد. تنها مورد نسبتاً دشوار مطابق شکل زیر است.
- **(خودمان)** با i مرتبه تزریق داریم:
  $m^2 + (i-1)k_1$ تا a و $m + (i-1)k_2$ تا b که $k_1, k_2 \neq 0$.
  اگر $i = 0$ باشد، خواهیم داشت:

$$i = 0 \Rightarrow (m - k_2)^2 \leq (m-1)^2 = m^2 - 2m + 1 < m^2 - k_1$$

پس $w_0 \notin L$.
همچنین اگر $k_1 = 0, k_2 \neq 0$ یا $k_1 \neq 0, k_2 = 0$، بازهم $w_0 \notin L$.
پس زبان L یک زبان مستقل از متن نیست.



مثال

نشان دهید که زبان $L = \{a^{n!} : n \geq 0\}$ مستقل از متن نیست.

این اثبات مشابه اثبات نامنظم بودن L است که در اینجا از بیان مجدد آن صرف نظر می‌کنیم.

۲-۱-۸ لم تزریق برای زبان‌های خطی

تعریف **زبان خطی.** زبان مستقل از متن L را خطی گوییم، اگر و فقط اگر یک گرامر مستقل از متن خطی مانند G وجود داشته باشد که $L = L(G)$.

هر زبان خطی یک زبان مستقل از متن است، اما عکس آن همواره صحیح نیست.

مثال

$\{a^n b^n : n \geq ۰\}$ یک زبان خطی است؛
زیرا گرامر خطی $S \to aSb \mid \lambda$ برای آن وجود دارد.

مثال

$\{w : n_a(w) = n_b(w)\}$ یک زبان خطی نیست؛ زیرا هیچ گرامر خطی برای آن وجود ندارد.

قضیه فرض کنید که L یک زبان خطی نامتناهی باشد، در این صورت عدد صحیح مثبت m وجود دارد که هر $w \in L$ با $|w| \geq m$ می‌تواند به صورت زیر تجزیه شود:

$$w = uvxyz$$

که در آن

$$|uvyz| \leq m, \qquad |vy| \geq ۱$$

به طوری که به ازای هر $i = ۰, ۱, ۲, \ldots$، $uv^ixy^iz \in L$ است.

◄ **اثبات.**
برای اثبات، دنباله‌ی اثبات اولین قضیه‌ی این فصل را پی می‌گیریم:
چون L خطی است، گرامر خطی G برای آن وجود دارد که به علاوه هیچ قاعده‌ی یکه یا λ ندارد (حذف این قواعد در صورت وجود خطی بودن گرامر را تغییر نمی‌دهد).
در درخت اشتقاق شکل ارایه شده در اثبات اولین قضیه‌ی این فصل می‌بینیم که:

- از آنجا که گرامر خطی است، متغیرها تنها می‌توانند در مسیر S به اولین A، از آن به دومین A و از آن به ... تا یک برگ قرار داشته باشند.
- از آنجا که تنها تعداد محدودی پایانه وجود دارد، u و z باید کران‌دار باشند.
 بر اساس همین بحث باید v و y هم کران‌دار باشند، پس:

$$|uvyz| \leq m$$

ادامه‌ی اثبات مشابه اثبات قضیه‌ی اول است.

کاربرد این لم تزریق، برای نشان دادن خطی نبودن یک زبان است.

مثال

نشان دهید که زبان $L = \{w : n_a(w) = n_b(w)\}$ خطی نیست.

فرض می‌کنیم که این زبان خطی باشد.
لم تزریق را برای رشته‌ی $w = a^m b^{۲m} a^m$ که $w \in L$ و $|w| \geq m$ است را در نظر می‌گیریم.
شرط $|uvyz| \leq m$ نشان می‌دهد که u, v, y, z باید حتماً شامل a باشد.
اگر این رشته را تزریق کنیم، داریم:

$$a^{m+k} b^{۲m} a^{m+l} \notin L, \quad k \geq ۱ \text{ or } l \geq ۱$$

پس زبان L نمی‌تواند خطی باشد.

۸-۲ خصوصیات بستاری و الگوریتم‌های تصمیم‌گیری برای زبان‌های مستقل از متن

۸-۲-۱ بستار زبان‌های مستقل از متن

قضیه خانواده‌ی زبان‌های مستقل از متن تحت اعمال اجتماع، الحاق و بستار ستاره‌ای بسته است.

◄ **اثبات.**
(امکان ساخت گرامر مستقل از متن برای حاصل اعمال فوق)
اگر L_1 و L_2 دو زبان مستقل از متن دلخواه باشند، آنگاه گرامرهای مستقل از متن زیر را می‌توان برای آنها ارایه کرد:

$$L_1 = L(G_1), \qquad G_1 = (V_1, T, S_1, P_1)$$
$$L_2 = L(G_2), \qquad G_2 = (V_2, T, S_2, P_2)$$
$$V_1 \cap V_2 = \varnothing, \qquad S \notin V_1 \cup V_2$$

مطابق جدول زیر، می‌توان برای اجتماع و الحاق این دو زبان و بستار ستاره‌ای هر یک از آنها یک گرامر مستقل از متن ارایه کرد (اثبات هر مورد بر اساس تعریف):

$L_1 \cup L_2 = L(G)$	$G = (V_1 \cup V_2 \cup \{S\}, T, S, P_1 \cup P_2 \cup \{S \to S_1 \mid S_2\})$
$L_1 L_2 = L(G)$	$G = (V_1 \cup V_2 \cup \{S\}, T, S, P_1 \cup P_2 \cup \{S \to S_1 S_2\})$
$L_1^* = L(G)$	$G = (V_1 \cup \{S\}, T, S, P_1 \cup \{S \to SS_1 \mid \lambda\}) \equiv$ $G = (V_1 \cup \{S\}, T, S, P_1 \cup \{S \to S_1 S \mid \lambda\})$

پس خانواده‌ی زبان‌های مستقل از متن تحت اعمال اجتماع، الحاق و بستار ستاره‌ای بسته است.

■

قضیه خانواده‌ی زبان‌های مستقل از متن تحت اعمال اشتراک و مکمل بسته نیست.

◄ **اثبات.**
(مثال نقض برای اشتراک و استفاده از قانون دمورگان برای مکمل.)
زبان‌های L_1 و L_2 را به صورت زیر در نظر می‌گیریم.

$$L_1 = \{a^n b^n c^m : n \geq 0, m \geq 0\}, \qquad L_2 = \{a^n b^m c^m : n \geq 0, m \geq 0\}$$

هر دو این زبان‌ها مستقل از متن هستند، زیرا برای L_1 گرامر مستقل از متن

$$\begin{aligned} S &\rightarrow S_۱S_۲ \\ S_۱ &\rightarrow aS_۱b \mid \lambda \\ S_۲ &\rightarrow cS_۲ \mid \lambda \end{aligned}$$

و برای $L_۲$ گرامر مستقل از متن

$$\begin{aligned} S &\rightarrow S_۱S_۲ \\ S_۱ &\rightarrow aS_۱ \mid \lambda \\ S_۲ &\rightarrow bS_۲c \mid \lambda \end{aligned}$$

وجود دارد. اما اشتراک آنها

$$L_۱ \cap L_۲ = \{a^n b^n c^n : n \geq ۰\}$$

دیگر مستقل از متن نیست. پس خانواده‌ی زبان‌های مستقل از متن تحت اشتراک بسته نیست.
داریم:

$$L_۱ \cap L_۲ = \overline{\overline{L_۱} \cup \overline{L_۲}}$$

با فرض مستقل از متن بودن $L_۱$ و $L_۲$ اگر زبان‌های مستقل از متن تحت مکمل بسته باشند، باید $\overline{L_۱}$ و $\overline{L_۲}$ مستقل از متن باشند و از آنجا که اجتماع دو زبان مستقل از متن، مستقل از متن خواهد بود، با فرض بسته بودن نسبت به مکمل، $L_۱ \cap L_۲$ باید مستقل از متن باشد. در صورتی که چنین چیزی درست نیست. پس خانواده‌ی زبان‌های مستقل از متن تحت مکمل بسته نیست.

■

قضیه خانواده‌ی زبان‌های مستقل از متن تحت عمل اشتراک منظم *(regular intersection)* بسته است: اگر $L_۱$ یک زبان مستقل از متن و $L_۲$ یک زبان منظم باشد، آنگاه $L_۱ \cap L_۲$ مستقل از متن است.

◄ **اثبات.**

فرض کنید که $M_۱$ یک NPDA باشد که زبان مستقل از متن $L_۱$ را می‌پذیرد

$$M_۱ = (Q, \Sigma, \Gamma, \delta_۱, q_۰, z, F_۱), \quad L_۱ = L(M_۱)$$

و $M_۲$ یک DFA باشد که زبان منظم $L_۲$ را می‌پذیرد

$$M_۱ = (P, \Sigma, \delta_۲, p_۰, F_۲), \quad L_۱ = L(M_۱)$$

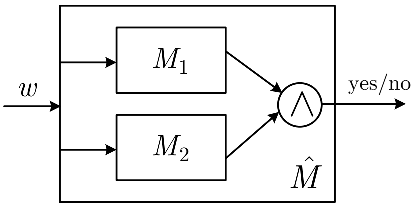
یک NPDAی $\hat{M}$ را به گونه‌ای می‌سازیم که $M_۱$ و $M_۲$ را به طور همزمان شبیه‌سازی کند:

$$\hat{M} = (\hat{Q}, \Sigma, \Gamma, \hat{\delta}, \hat{q}_۰, z, \hat{F})$$

(هرگاه یک نماد از رشته‌ی ورودی خوانده می‌شود، $\hat{M}$ به طور موازی حرکت‌های $M_۱$ و $M_۲$ را اجرا می‌نماید.)

فرض می‌کنیم:

$$\begin{aligned}
\hat{Q} &= Q \times P \\
\hat{q}_۰ &= (q_۰, p_۰) \\
\hat{F} &= F_۱ \times F_۲ \\
\hat{\delta} &: ((q_k, p_l), x) \in \hat{\delta}((q_i, p_j), a, b) \Leftrightarrow (q_k, x) \in \delta_۱(q_i, a, b) \wedge \delta_۲(p_j, a) = p_l \\
&: a = \lambda \Rightarrow p_j = p_l
\end{aligned}$$



حالات $\hat{M}$ برچسب‌گذاری شده با (q_i, p_j) بیانگر حالات متناظر در $M_۱$ و $M_۲$ است که پس از خواندن یک رشته‌ی ورودی یکسان در آن خواهند بود.
می‌توان با استقرا نشان داد که:

$$((q_۰, p_۰), w, z) \vdash^* ((q_r, p_s), x), q_r \in F_۱, p_s \in F_۲ \Leftrightarrow (q_۰, w, z) \vdash^*_{M_۱} (q_r, x) \wedge \delta^*_۲(p_۰, w) = p_s$$

بنابراین یک رشته پذیرفته می‌شود اگر و فقط اگر هم توسط $M_۱$ و هم توسط $M_۲$ پذیرفته شود، یعنی $L(M_۱) \cap L(M_۲) = L_۱ \cap L_۲$.
پس برای اشتراک یک زبان مستقل از متن و یک زبان منظم، می‌توان یک NPDA ارایه کرد که آن را بپذیرد. پس با یک زبان مستقل از متن سر و کار داریم.

مثال

نشان دهید که زبان $L = \{a^n b^n : n \geq ۰, n \neq ۱۰۰\}$ مستقل از متن است.

$$L = \underbrace{\underbrace{\overline{L(a^{۱۰۰} b^{۱۰۰})}}_{\text{منظم}} \cap \underbrace{\{a^n b^n : n \geq ۰\}}_{\text{مستقل از متن}}}_{\text{مستقل از متن}}$$

مثال

زبان $L = \{w \in \{a, b, c\}^* : n_a(w) = n_b(w) = n_c(w)\}$ مستقل از متن نیست.

زیرا طبق قضیه‌ی بستار زبان‌های مستقل از متن تحت عمل اشتراک منظم، باید اشتراک L با زبان منظم $L(a^*b^*c^*)$ مستقل از متن باشد:

$$L \cap L(a^*b^*c^*) = \{a^n b^n c^n : n \geq ۰\}$$

اما این زبان مستقل از متن نیست.

قضیه خانواده‌ی زبان‌های مستقل از متن تحت عمل تبدیل همریختی بسته است.

قضیه خانواده‌ی زبان‌های مستقل از متن تحت عمل معکوس *(reverse)* بسته است.

قضیه خانواده‌ی زبان‌های مستقل از متن تحت عمل خارج قسمت راست منظم *(regular right quotient)* بسته است: اگر $L_۱$ یک زبان مستقل از متن و $L_۲$ یک زبان منظم باشد، آن‌گاه $L_۱/L_۲$ مستقل از متن است.

۸-۲-۲ مسائل تصمیم‌گیری برای زبان‌های مستقل از متن

قضیه اگر گرامر مستقل از متن $G = (V, T, S, P)$ در دست باشد، الگوریتمی برای تصمیم‌گیری در مورد اینکه آیا $L(G)$ تهی است یا خیر، وجود دارد.

◄ **اثبات.**
اگر $\lambda \notin L(G)$ از الگوریتم حذف قواعد بی‌فایده استفاده می‌کنیم. اگر S حذف شد، $L(G)$ تهی است، وگرنه حداقل شامل یک عنصر است.

قضیه اگر گرامر مستقل از متن $G = (V, T, S, P)$ در دست باشد، الگوریتمی برای تصمیم‌گیری در مورد اینکه آیا $L(G)$ متناهی است یا خیر، وجود دارد.

◄ **اثبات.**

از گراف وابستگی گرامر استفاده می‌کنیم. اگر پس از ساده‌سازی گرامر، در گراف وابستگی از یک متغیر مفید به خودش یک دور وجود داشت (متغیر تکرارشونده)، زبان آن گرامر نامتناهی و در غیر این صورت متناهی است.

■

قضیه برای زبان‌های مستقل از متن، برای پاسخ به پرسش مسئله‌ی عضویت $(w \in L)$ الگوریتم وجود دارد.

◄ **اثبات.**
از آنجا که برای هرگرامر مستقل از متن (بدون λ)، معادلی در فرم چامسکی وجود دارد، گرامر L را به فرم چامسکی تبدیل می‌کنیم و الگوریتم CYK برای عضویت را به کار می‌بریم.

■

قضیه برای تصمیم‌گیری در مورد معادل بودن دو گرامر مستقل از متن، الگوریتمی وجود ندارد.

قضیه برای تصمیم‌گیری در مورد اینکه آیا زبان گرامر G، $L(G)$ مساوی با Σ^* است یا خیر، الگوریتمی وجود ندارد.

قضیه برای تصمیم‌گیری در مورد مبهم بودن یک گرامر مستقل از متن، الگوریتمی وجود ندارد.

مراجع

[1] P. Linz, **An Introduction to Formal Languages and Automata**, 5th Ed., Jones and Bartletts, 2012.

[2] M. Sipser, **Introduction to the Theory of Computation**, 3rd Ed., Cengage Learning, 2013.