دانشکده‌ی علوم ریاضی

نظریه‌ی زبان‌ها و اتوماتا ۲۳ آبان ۱۳۹۱

جلسه‌ی ۱۸: حذف قوانین تولید ε و قوانین یکه

مدرّس: دکتر شهرام خزائی *نگارندگان: علی پورشاه‌آبادی و محمدصالح یاشیخ‌اکبری*

۱ مقدمه

در جلسه‌ی قبل متغیر **بی‌استفاده** [۱] و روش حذف آن‌ها معرفی شد. در ادامه کار برای بدست آوردن فرم نرمال چامسکی یک گرامر این جلسه در مورد حذف قوانین تولید ε و حذف قوانین یکه بحث می‌کنیم.

۲ حذف قوانین تولید ε

تعریف ۱ *به قوانینی به شکل زیر **قانون تولید** ε میگویند.*

$$A \rightarrow \varepsilon$$

برای به دست آوردن فرم نرمال چامسکی ما باید قوانین تولید ε را از قوانین تولید زبان L حذف کنیم. بعد از حذف قوانین تولید ε از قوانین گرامر زبان L، زبان حاصل باید به صورت $L - \{\varepsilon\}$ بشود.

تعریف ۲ *متغیر A را **تهی‌ساز** [۲] گوییم اگر $A \overset{*}{\Rightarrow} \varepsilon$ (یعنی بتوانیم ε را طی چند مرحله اشتقاق از متغیر A به دست بیاوریم.)*

۱.۲ الگوریتم کشف متغیرهای تهی‌ساز

با استفاده از الگوریتم بازگشتی زیر همه‌ی متغیرهای تهی‌ساز پیدا می‌شوند:
حالت پایه: اگر $A \rightarrow \varepsilon$ یک قانون تولید باشد A یک متغیر تهی‌ساز است.

حالت استقرا: اگر $A \rightarrow \alpha$ که $(\alpha \in V^*)$ یک قانون تولید باشد و همه ی متغیرهای موجود در α تهی ساز باشند ، آنگاه A نیز تهی‌ساز است.

مثال ۱ *می‌خواهیم متغیرهای تهی‌ساز گرامر زیر را پیدا کنیم:*

$$\begin{array}{lcl} S & \rightarrow & ABC \\ A & \rightarrow & aA|a|\varepsilon \\ B & \rightarrow & bB|\varepsilon \\ C & \rightarrow & \varepsilon \end{array}$$

[۱] useless
[۲] nullable

حالت پایه: به طور مستقیم متغیرهای A و B و C، رشته ε را تولید می‌کنند.
حالت استقرا:

$$S \rightarrow ABC$$

همه‌ی متغیرهای سمت راست این قانون تولید متغیرهای تهی‌ساز هستند، پس S نیز یک متغیر تهی‌ساز است.
بنابراین S, A, B, C مجموعه همه متغیرهای تهی‌ساز می‌باشد.

۲.۲ الگوریتم حذف قوانین تولید ε

ابتدا تمام متغیرهای تهی‌ساز را پیدا می‌کنیم.
اگر قانون تولیدی به شکل $A \rightarrow X_1X_2...X_n$ وجود داشته باشد که دارای $m \leqslant n$ متغیر تهی‌ساز می‌باشد، باید این قانون تولید را با 2^m قانون جایگزین کنیم که در هر قانون جدید یکی از 2^m زیررشته‌ی تهی‌ساز وجود ندارد.

نکته : در حالتی که $m = n$ است ما زیر رشته‌ایی که شامل تمام متغییرهای تهی‌ساز است را حذف نمی‌کنیم.

به عنوان مثال اگر ما قانون تولیدی به شکل $A \rightarrow BCD$ داشته باشیم که در آن متغییر های A و B تهی ساز باشند آنگاه این قانون تولید را باید به صورت چهار (2^2) قانون تولید جدید به صورت $A \rightarrow BCD|CD|BD|D$ نوشت. ولی اگر قانون تولید ما به صورت $A \rightarrow BC$ بود آنگاه قوانین تولید جدید ما به صورت $A \rightarrow BC|C|B$ می‌شود.

اگر این روش را استفاده کنیم زبانی که یک گرامر مستقل از متن می‌پذیرد تغییر نمی‌کند.

قضیه ۱ *اگر G یک گرامر مستقل از متن و G' گرامری باشد که پس از اعمال الگوریتم حذف قوانین ε به دست می‌آید، داریم:*

$$L(G') = L(G) - \{\varepsilon\}$$

مثال ۲ *می‌خواهیم قوانین تولید ε را از گرامر مثال قبل حذف کنیم. به طوری که گرامر جدیدی که حاصل می شود همان زبان به غیر از ε را بپذیرد.*
حل :

$$\begin{array}{lcl} S & \rightarrow & ABC|BC|AC|AB|A|B|C \\ A & \rightarrow & aA|a \\ B & \rightarrow & bB|b \\ D & \rightarrow & abAa|aba \end{array}$$

۳ قوانین یکه

تعریف ۳ *یک قانون تولید $A \rightarrow B$ که $A, B \in V$ **قانون یکه**[3] نامیده میشود.*

تعریف ۴ *زوج (A, B) را که $A, B \in V$ یک زوج یکه می‌گوییم اگر $A \overset{*}{\Rightarrow} B$*

۱.۳ الگوریتم کشف زوج های یکه

با استفاده از الگوریتم بازگشتی زیر زوج‌های یکه را پیدا می‌کنیم. پایه: برای هر متغیر $A \in V$، زوج (A, A) یکه است.
استقرا: اگر (A, B) یک زوج یکه باشد و قانون تولید یکه $B \rightarrow C$ را داشته باشیم، آنگاه (A, C) نیز یک زوج یکه است.

مثال ۳ *در گرامر زیر زوج های یکه را مشخص کنید.*

$$\begin{array}{lcl} I & \rightarrow & a|b|Ia|Ib|I0|I1 \\ F & \rightarrow & I|(E) \\ T & \rightarrow & F|T*F \\ E & \rightarrow & T|E+F \end{array}$$

[3] Unit production

حل:
حالت پایه: همه زوج‌های زیر بنابر تعریف زوج یکه هستند.

$$(I,I),(F,F),(T,T),(E,E)$$

حالت استقرا: با توجه به زوج‌های یکه‌ایی که تا حالا بدست آورده‌ایم و قوانین تولید داده شده، زوج‌های زیر نیر زوج یکه به حساب می‌آیند:

$$(F,I),(T,F),(E,T)$$

دوباره با توجه به زوج‌های یکه

$$(F,I),(T,F),(E,T)$$

که بدست آوردیم و قوانین تولید داده شده زوج‌های زیر نیز زوج یکه بحساب می‌آیند:

$$(T,I),(E,F)$$

دوباره با توجه به زوج‌های یکه

$$(T,I),(E,F)$$

که بدست آوردیم و قوانین تولید داده شده زوج زیر نیز زوج یکه بحساب می آیند:

$$(E,I)$$

پس در کل زوج های یکه برای این قوانین روج‌های زیر هستند:

$$\{(I,I),(F,F),(T,T),(E,E),(F,I),(T,F),(E,T),(T,I),(E,F),(E,I)\}$$

قضیه ۲ *الگوریتم کشف زوج‌های یکه همه زوج‌های یکه را پیدا می‌کند و هر زوج را که پیدا کند حتما زوج یکه است. یعنی اگر زوج (A,B) توسط این الگوریتم کشف شود آنگاه حتما زوج (A,B) یک زوج یکه است. و همچنین اگر زوج (A,B) یک زوج یکه باشد آنگاه این الگوریتم حتما زوج (A,B) را کشف می کند.*

در ادامه کار (در مسیر بدست آوردن فرم نرمال چامسکی) باید زوج‌های یکه را از گرامر حذف کنیم ولی باید طوری زوج‌های یکه را از گرامر حذف کنیم که زبان گرامر جدیدی که بدست می‌آید تغییر نکند.

۲.۳ الگوریتم حذف قواعد یکه

به ازای هر قانون تولید $B \to \alpha$ که $\alpha \notin V$ (یعنی $B \to \alpha$ یکه نیست) اگر (A,B) یک زوج یکه باشد آنگاه $A \to \alpha$ را به عنوان یک قانون تولید در نظر بگیرید.

قضیه ۳ *پس از اعمال الگوریتم حذف قواعد یکه گرامر جدیدی که حاصل می‌شود قانون یکه ندارد و زبان گرامر جدید همان زبان گرامر قبلی است .*

مثال ۴ *زوج‌های یکه را از قوانین تولید مثال ۳ حذف کنید:*
حل:
برای این کار به صورت زیر عمل می کنیم :

زوج های یکه	قوانین تولید جدید
(E,E)	$E \to E+F$
(E,T)	$E \to T*F$
(E,F)	$E \to (E)$
(E,I)	$E \to a|b|Ia|Ib|I0|I1$
(T,T)	$T \to T*F$
(T,F)	$T \to (E)$
(T,I)	$T \to a|b|Ia|Ib|I0|I1$
(F,F)	$F \to (E)$
(F,I)	$F \to a|b|Ia|Ib|I0|I1$
(I,I)	$I \to a|b|Ia|Ib|I0|I1$

پس با توجه به جدول بالا قوانین تولید جدید به صورت زیر می شوند:

$$\begin{array}{lcl} E & \to & E+F|T*F|(E)|a|b|Ia|Ib|I0|I1 \\ T & \to & T*F|(E)|a|b|Ia|Ib|I0|I1 \\ F & \to & (E)|a|b|Ia|Ib|I0|I1 \\ I & \to & a|b|Ia|Ib|I0|I1 \end{array}$$

۳.۳ بدست آوردن فرم نرمال چامسکی برای یک گرامر

با توجه به مطالبی که گفته شد داریم:

قضیه ۴ *اگر L یک زبان مستقل از متن باشد یک گرامر برای $L-\{\varepsilon\}$ وجود دارد که:*

- *دارای متغیر بیاستفاده نیست .*
- *داری قانون تولید ε نیست.*
- *دارای قانون تولید یکه نیست.*

برهان. از روی گرامر G برای زبان L گرامر G' را به صورت زیر می سازیم (ترتیب این عملیات خیلی مهم است):

۱. حذف قوانین تولید ε

۲. حذف قوانین یکه

۳. حذف متغیرهای غیر مولد

۴. حذف متغیرهای غیر قابل دسترس

می‌توان نشان داد که گرامر G' ویژگی مورد نظر قضیه را دارد.

■

پس از اعمال این چهار عملیات بر روی یک گرامر قانون های تولید جدید به این صورت میشوند که بدنه هر قانون تولید یا یک نماد تکی پایانه است یا رشته‌ای با طول بزرگتر یا مساوی ۲ از پایانه‌ها و متغیرها است.

تعریف ۵ *می‌گوییم $G=(V,T,P,S)$ به فرم نرمال چامسکی است اگر هر قانون تولید آن به یکی از دو صورت زیر باشد:*

۱. $A \to a$ که $A \in V$ و $a \in T$

۲. $A \to BC$ که $A, B, C \in V$

قضیه ۵ *اگر L یک زبان مستقل از متن باشد آنگاه یک گرامر به فرم نرمال چامسکی برای $L - \{\varepsilon\}$ وجود دارد.*

مثال ۵ *گرامر زیر را به فرم نرمال چامسکی تبدیل کنید:*

$$\begin{array}{lcl} S & \to & ABa \\ A & \to & aab \\ B & \to & Ac \end{array}$$

برای تبدیل به فرم نرمال چامسکی ابتدا متغیرهایی که یک نماد تکی پایانه داشته باشند را تولید میکنیم.
متغیرهایی جدیدی برای پایانه‌ها به صورت زیر تعریف می‌کنیم:

$$\begin{array}{lcl} C & \to & a \\ D & \to & b \\ E & \to & c \end{array}$$

برای شکستن متغیرهای S و A متغیرهای جدیدی نیز به صورت زیر تعریف می‌کنیم:

$$\begin{array}{lcl} F & \to & BC \\ G & \to & CD \end{array}$$

پس گرامر زیر به دست می‌آید که به فرم نرمال چامسکی است:

$$\begin{array}{lcl} S & \to & AF \\ A & \to & CG \\ B & \to & AE \\ C & \to & a \\ D & \to & b \\ E & \to & c \\ F & \to & BC \\ G & \to & CD \end{array}$$