논·문 14-2-2

단순한 LL 커버링 변환
(Simple LL Covering Transformation)

이경옥
한신대학교 컴퓨터정보통신학부 정보통신학전공
golee@hucc.hanshin.ac.kr

최광무
한국과학기술원 전자전산학과 전산학전공
choe@cs.kaist.ac.kr

요 약

LR 문법에 대한 LL 커버링 변환에 대한 기존의 방법들은 복잡한 정형식을 필요로 하기에 그 근본 아이디어를 이해하기 어렵다. 본 논문에선 LL 커버링 변환의 근본 아이디어를 설명한 후에, 간단하며 직관적인 LL 커버링 변환을 제시한다.

1. 서론

LR 문법의 LL 문법으로의 커버링 성질을 만족하는 변환은 이론적, 실용적인 관심의 대상이 되어왔다 [H74a, H74b, SU79, L00]. 한편 기존의 방법들은 복잡한 정형식들에 근거했기에, 그 근본 아이디어를 이해하기 어렵다.

본 논문에선 LL 커버링 변환의 근본 아이디어를 설명한 후에 간단하면서도 이해하기 쉬운 커버링 변환을 제안한다.

2. 기본 정의

표기법과 기본 정의들은 [AU72,SS90]을 따르며 독자의 편의를 위해서 몇 가지 정의들을 다시 언급한다.

심볼 G는 임의의 고정된 $G=(N,\Sigma,P,S)$ ($V=N\cup\Sigma$)를 나타낸다. 심볼 k는 임의의 고정된 음수가 아닌 정수를 나타낸다. N에 속하지 않는 S'에 대해서 $G'=(N\cup\{S'\},\Sigma,P\cup\{S'\rightarrow S\$^{k}\},S')$는 G의 **확장된(augmented) 문법**이다. G는

S'에 의해 확장된 문법을 가정하고 또한 **리듀스된 (reduced) 문법** [SS90]을 가정한다. 그리고 G는 어떤 $k(k\geq 1)$에 대한 LR(k) 문법임을 가정한다. α를 스트링이라 하자. $k:\alpha$는 $|\alpha|\leq k$이면 α와 동일하고 그 외의 경우엔 α의 처음 k개의 심볼들과 동일하다. 이때 $FIRST_k(\alpha)=\{k:x|\alpha=>^{*}x, x\in\Sigma^{*}\}$이다.

3. 변환의 원리

이 장에서는 LR 문법의 특성을 이용한 LL 커버링 문법을 얻는 방법에 대한 근본 아이디어를 설명한다. 주어진 문법 G에 대한 변환 문법 $T(G)$의 규칙들이 다음과 같이 생성되면 $T(G)$는 G의 커버링 문법이다.

Type 1. $[S',\alpha]\rightarrow a[S',\alpha a]$

이때

(조건1) $S'=>^{*}_{rm}\beta Bw=>_{rm}\beta\gamma a\delta w(\alpha=\beta\gamma)$

가 G상에 존재해야한다.

Type 2. $[S',\alpha]\rightarrow[S',\beta B]$

이때

(조건2) $S'=>^{*}_{rm}\beta Bw=>_{rm}\beta\gamma w(\alpha=\beta\gamma)$가 G상에 존재해야한다.

Type 3. $[S',S']\rightarrow\varepsilon$

h를 $T(G)$상의 규칙들로부터 G상의 규칙들로의 함수로 다음과 같이 정의하자.

$r=[S',\beta\gamma]\rightarrow[S',\beta B]$이면 $h(r)=B\rightarrow\gamma$이고 그 밖의 경우엔 $h(r)=\varepsilon$이다. 이때 $T(G)$는 G를 h의 관점에서 좌측에서 우측으로 커버(left-to-right cover)함을 알 수 있다.

위와 같이 정의되는 $T(G)$의 넌터미널들의 집합은 유한성의 조건을 만족해야한다. 일반적으로 어떤 문법 G의 바이어블 프리픽스는 무한하기에 $T(G)$의 문법으로 G에 대한 커버링 변환을 얻을 수 있는 문법의 범위는 매우 작다. G가 LR(k) 문법임을 가정했기에 Type 1 규칙과 Type 2 규칙간의 produce/produce 충돌이 발생하지 않고, Type 2 규칙과 Type 2 규칙간의 produce/produce 충돌이 발생하지 않는다. 이와 같은 원리로 $T(G)$는 LL(k) 문법임을 알 수 있다.

다음에선 $T(G)$를 구성하는 규칙에 새로운 Type의 규칙을 추가하여, $T(G)$로 커버링할 수 있는 문법의 범위를 넓히려고 한다.

4. Π^s 관계를 이용한 변환

이 장에선 새로운 Type의 규칙의 생성을 위한 Π^s 관계를 정의한다. 이에 바탕을 둔 확장된 LL 커버링 변환 문법을 제안한다.

정의(Π^s 관계)

$(A,\alpha)\ \Pi^s\ B$가 성립하기 위해서는 다음이 성립한다. 즉, 다음의

(유도과정 1)

$$A=>^{+}_{rm}\alpha w,\ k:w\in FIRST_k(B)$$

이 존재할 때마다 (유도과정 1)은

(유도과정 2)

$$A=>^{+}_{rm}\alpha Bz=>^{*}_{rm}\alpha yz,\ w=yz$$

의 형태이다. □

(유도과정 2)는 (유도과정 1)에서의 B에 대한 예상을 의미한다.

LL 성질을 위해서 Π^s 관계의 부분 관계 Π^s_{LL}을 정의한다.

정의 (Π^s_{LL} 관계) Π^s_{LL} 관계는 Π^s의 부분

관계로서 다음의 조건을 만족한다. 즉 $(A, \alpha)\ \Pi^s_{LL}\ B$, $(A, \alpha) \Pi^s_{LL}\ C$이면 $FIRST_k(B) \cap FIRST_k(C) = \varnothing$이다. □

다음에선 Π^s_{LL}의 부분 관계 $\widehat{\Pi^s}$에 기반한 변환을 제시한다.

알 고 리 즘 ($T2(G, \widehat{\Pi^s}) = (N2, \Sigma, F2, S2)$의 생성)

입력: $G, \widehat{\Pi^s}(\widehat{\Pi^s} \subseteq \Pi^s_{LL})$

출력: 이 알고리즘이 올바로 끝나면 $T2(G, \widehat{\Pi^s})$를 생성한다.

방법:

$F2 = \varnothing, S2 = [S', \varepsilon], N2 = \{S2\}$으로 초기화한다.

repeat

각 $[A, \alpha] \in N2$에 대해서 다음을 수행한다.

$X = \{x \in FIRST_k(B) | (A, \alpha)\ \ \widehat{\Pi^s}\ B\}$이라고 한다.

Type1. $F2 = F2 \cup \{[A, \alpha] \rightarrow a[A, \alpha a]\}$
이때

$$A =>^*_{\mathrm{rm}} \beta Bw =>_{\mathrm{rm}} \beta\gamma a\delta w (\alpha = \beta\gamma)$$

$=>^*_{\mathrm{rm}} \beta\gamma azw$가 G상에 존재하고

$$\{k: azw | A =>^*_{\mathrm{rm}} \beta Bw =>_{\mathrm{rm}} \beta\gamma a\delta w (\alpha = \beta\gamma) =>^*_{\mathrm{rm}} \beta\gamma azw\} \cap X = \varnothing$$이어야 한다.

Type2. $F2 = F2 \cup \{[A, \alpha] \rightarrow [A, \beta C]\}$
이때

$A =>^*_{\mathrm{rm}} \beta Bw =>_{\mathrm{rm}} \beta\gamma w (\alpha = \beta\gamma)$가 G상에 존재하고

$$\{k: w | A =>^*_{\mathrm{rm}} \beta Bw =>_{\mathrm{rm}} \beta\gamma w (\alpha = \beta\gamma)\} \cap X = \varnothing$$이어야 한다.

Type3. $F2 = F2 \cup \{[A, A] \rightarrow \varepsilon\}$

Type4. $F2 = F2 \cup \{[A, \alpha] \rightarrow [B, \varepsilon][A, \alpha B]\}$
이때

$(A, \alpha)\ \ \widehat{\Pi^s}\ B$이 성립해야한다.

until ($N2$집합이 변하지 않는다.) □

$h2$를 $T2(G, \widehat{\Pi^s})$의 규칙들로부터 G의 규칙들로의 함수로 다음과 같이 정의하자.
$r = [A, \beta\gamma] \rightarrow [A, \beta B]$이면 $h2(r) = B \rightarrow \gamma$이고 그 밖의 경우엔 $h2(r) = \varepsilon$이다.

보조정리 1. $[A, \alpha] \in N2$라 하자. $A =>^{\pi}_{\mathrm{rm}} \alpha x$인 π가 G상에 존재하면 $h2(\pi_T) = \pi^R$이고 $[A, \alpha] =>^{\pi_T}_{lm} x$인 π_T가 $T2(G, \widehat{\Pi^s})$상에 존재한다.

(증명) $|\pi|$에 대한 인덕션에 의해 증명한다. 초기 단계로서 $|\pi| = 1$이라고 하자. 이때 $A \rightarrow \alpha x \in P$이고, $[A, \alpha] =>^{\pi_T}_{lm} x[A, \alpha x]$인 π_T가 존재하고, $[A, \alpha x] =>^{r_T}_{lm} [A, A] =>_{lm} \varepsilon$인 r_T가 $T2(G, \widehat{\Pi^s})$상에 존재한다. 여기서 $h2(\pi_T r_T) = h2(\pi_T) h(r_T) = \varepsilon \cdot r = A \rightarrow \alpha x$이다. 따라서 초기 단계에 대해선 본 보조 정리가 성립한다.

인덕션 가정으로 $|\pi| \leq l (l > 1)$일 때 이 보조 정리가 성립한다고 가정하자. 그리고 $|\pi| > l$이라고 가정하자.

(i) $(A, \alpha)\ \widehat{\Pi^s} B$가 성립한다면 π를 $A =>^{\pi^1}_{\mathrm{rm}} \alpha Bz$인 π^1과 $B =>^{\pi^2}_{\mathrm{rm}} y$인 π^2로 나눌 수 있고 $[A, \alpha] \rightarrow [B, \varepsilon][A, \alpha B]$인 규칙

r_T가 존재함을 안다. π^1과 π^2에 대해서 인덕션 가정으로 $[A,\alpha] \Rightarrow^{\pi^1_T}_{lm} z$, $[B,\varepsilon] \Rightarrow^{\pi^2_T}_{lm} y$인 π^1_T와 π^2_T가 $T2(G,\widehat{\Pi^s})$상에 존재한다. 여기서 $h2(\pi^1_T) = (\pi^1)^R$이고 $h2(\pi^2_T) = (\pi^2)^R$이다. 따라서 $[A,\alpha] \Rightarrow^{\pi_T}_{lm} yz(=x)$인 $\pi_T(= r_T\pi^2_T\pi^1_T)$가 $T2(G,\widehat{\Pi^s})$상에 존재한다. 여기서 $h2(r_T\pi^2_T\pi^1_T) = \varepsilon \cdot h2(\pi^2_T)\,h2(\pi^1_T) = (\pi^2)^R(\pi^1)^R = (\pi^1\pi^2)^R = \pi^R$이 성립한다.

(ii) 그 밖의 경우에 대해서, π가 $A \Rightarrow^{\pi'}_{rm} \beta Bz$인 π'과 $\beta Bz \Rightarrow^{r}_{rm} \beta\gamma yz(\beta\gamma = \alpha, x = yz)$인 r로 구성되어 있다고 가정하자. 그리고 $y = a_1 \cdots a_n(n \geq 0)$이라고 하자.

(ii-a) $(A, \alpha a_1 \cdots a_i)\ \widehat{\Pi^s}\ C$인 $i(0 \leq i \leq n)$가 존재한다면, π는 $A \Rightarrow^{\pi^1}_{rm} \alpha a_1 \cdots a_i C z_2$인 π^1과 $C \Rightarrow^{\pi^2}_{rm} a_{i+1} \cdots a_n z_1 (z = z_1 z_2,\ \pi = \pi^1\pi^2)$인 π^2로 나눌 수 있다. 이때 $[A,\alpha]$에 Type 1의 규칙들을 반복적으로 적용하여 $[A,\alpha] \Rightarrow^{\pi^3_T}_{lm} a_1 \cdots a_i [A, \alpha a_1 \cdots a_i]$가 $T2(G,\widehat{\Pi^s})$상에 존재한다. 또한 $[A, \alpha a_1 \cdots a_i] \Rightarrow^{r_T}_{lm} [C,\varepsilon][A, \alpha a_1 \cdots a_i C]$인 r_T가 $T2(G,\widehat{\Pi^s})$상에 존재한다. 한편 π^2와 π^1에 인덕션 가정을 적용하여, $[C,\varepsilon] \Rightarrow^{\pi^2_T}_{lm} a_{i+1} \cdots a_n z_1$인 π^2_T와 $[A, \alpha a_1 \cdots a_i C] \Rightarrow^{\pi^1_T}_{lm} z_2$인 π^1_T가 존재한다. 여기서 $h2(\pi^2_T) = (\pi^2)^R$와 $h2(\pi^1_T) = (\pi^1)^R$이다. 따라서 $[A,\alpha] \Rightarrow^{\pi_T}_{lm} a_1 \cdots a_i a_{i+1} \cdots a_n z_1 z_2$인 $\pi_T(= \pi^3_T r_T \pi^2_T \pi^1_T)$가 존재한다. (여기서 $x = a_1 \cdots a_i a_{i+1} \cdots a_n$이다.) 그리고, $h2(\pi^3_T r_T \pi^2_T \pi^1_T) = \pi^R$이다.

(ii-2) 그 밖의 경우에 $[A,\alpha] \Rightarrow^{\pi^1_T}_{lm} a_1 \cdots a_n [A, \alpha a_1 \cdots a_n]$인 Type 1의 규칙으로 구성된 π^1_T가 존재한다. $T2(G,\widehat{\Pi^s})$의 생성방법에 의해 $[A, \alpha a_1 \cdots a_n] \Rightarrow^{r_T}_{lm} [A, \beta B]$인 r_T가 존재한다. 한편 π'에 인덕션 가정을 적용하여 $[A, \beta B] \Rightarrow^{\pi_T'}_{lm} z[A, A]$인 π_T'가 $T2(G,\widehat{\Pi^s})$상에 존재한다. 따라서 $[A,\alpha] \Rightarrow^{\pi_T}_{lm} a_1 \cdots a_n z\,[A, A]$인 π_T가 $T2(G,\widehat{\Pi^s})$상에 존재한다. 여기서 $x = a_1 \cdots a_n z$이고 $h2(\pi_T) = h2(\pi^1_T r_T \pi'_T) = \varepsilon \cdot h2(\pi_T) h(\pi_T') = (\pi' \cdot r)^R = \pi^R$이다. □

보조정리 2. $[A,\alpha] \Rightarrow^{\pi_T}_{lm} x$인 π_T가 $T2(G,\widehat{\Pi^s})$상에 존재하면 $A \Rightarrow^{\pi}_{rm} \alpha x$인 π가 G상에 존재한다. □

보조정리 1와 유사하게 π_T의 길이에 대한 인덕션으로 증명할 수 있다.

보조정리 1과 보조정리 2을 이용하여 다음의 정리 1을 증명할 수 있다.

정리 1. $T2(G,\widehat{\Pi^s})$는 $h2$의 관점에서 G를 좌측에서 우측으로 커버한다. □

G의 LR(k) 성질과 각 규칙 생성시의 Type 4 규칙과의 조건을 이용하여 다음의 정리를 얻을 수 있다. 자세한 증명은 생략한다.

정리 2. $T2(G, \widehat{\Pi^s})$는 LL(k) 문법이다. □

참고문헌

[AU72] Aho, A. V. and Ullman, J. D.:The Theory of Parsing, Translation and Compiling, vols.1 2.Englewood Cliffs, NJ:Prentice-Hall 1972, 1973

[H74a] Hammer, M.: A New Grammatical Transformation into deterministic top down form. MIT, Mass., Project MAC Technical Report TR-119, 1974

[H74b] Hammer, M.: A New Grammatical Transformation into LL(k) form. Proc. of Sixth Annual ACM Symposium on Theory of Computing, 266-275 (1974)

[L00] Lee, Gyung-Ok: On LL-to-LR Covering Grammars, Ph.D. Thesis, (2000)

[SS90] Sippu, S. and Soisalon-Soininen, E.:Parsing Theory, vol I, II, Springer-Verlag Berlin Heidelberg, 1990

[SU79] Soisalon-Soininen, E. and Ukkonen, E.: A Method for Transforming Grammars into LL(k) Form. Acta Informatica 12, 338-369 (1979)

이경옥

1986년3월~1990년2월 서강대학교 전자계산학과(학사).

1990년2월~1992년8월 한국과학기술원 전산학과(석사).

1992년9월~2000년2월 한국과학기술원 전산학과(박사).

2000년 3월~2000년 8월 한국과학기술원 전자전산학과 박사 후 연수연구원.

2000년 9월~현재 한신대학교 컴퓨터 및 정보통신학부 정보통신학전공 조교수.

관심분야는 컴파일러, 형식언어와 오토마타, 프로그래밍언어론

최광무

정보과학회지 제 26권 제 7호 (B) 참조