計算論 A / 2015年度

第7章 文脈自由言語の性質 (1/3)

文脈自由言語の標準形

角川　裕次

もくじ

第7章

＊＊＊ 本日の重要概念 ＊＊＊

チョムスキー標準形, ε-規則の除去

φ(・ω・) メモメモ

7.1 節での最終目標

任意の文脈自由言語 L:
L を生成するチョムスキー標準形の
文脈自由文法が存在

チョムスキー標準形とは

- 生成規則を以下の形に限定した文脈自由文法
  $A \to BC$, または
  $A \to a$
- ただし A, B, C は変数, a は終端記号

文脈自由言語に関する証明が見通し良くできて便利

7.1 文脈自由言語の標準形

文脈自由文法の等価変換

生成する言語は変えず文法を単純にする変換3種

- 無用な記号の除去
- ε-規則の除去
- 単位規則の除去

後述のチョムスキー標準形への変換のときに使う

7.1.1 無用な記号の除去

有用な記号

文法$G = (V, T, P, S)$の有用な記号$X (\in V \cup T)$

- ある終端記号列の導出でXが出現する場合をいう
- 即ち, $\exists w \in L(G) \ : \ S \overset{*}{\underset{G}{\Rightarrow}} \alpha X \beta \overset{*}{\underset{G}{\Rightarrow}} w$

無用な記号

文法$G = (V, T, P, S)$の無用な記号$X (\in V \cup T)$

- X が有用でない場合をいう
- 即ち, どの終端記号列の導出においてもXは出現しない

無用な記号は全く使われない

無用な記号を除去しても生成される言語は不変

- 生成規則は変更不要
- 文法はより単純になる

記号Xが有用となる2つの条件

Xは生成的: $X \overset{*}{\underset{G}{\Rightarrow}} w \ (\in T^*)$ となる場合

- もしそうでなければ...
  X からは終端記号列が全く得られない
  語の生成に役立ってない

Xは到達可能: $S \overset{*}{\underset{G}{\Rightarrow}} \alpha X \beta$ となる場合

- もしそうでなければ...
  そもそも X にたどり着くことがない
  語の生成に関して X には出番が回ってこない

X は有用: 上記2条件がともに満たされる場合

X は無用: それ以外の場合

有用な記号の求め方の手順

手順

1. 生成的でない記号を除去
2. 到達可能でない記号を除去

✔ この順序で正しい理由は?
逆の順序で行なったらどうなる?
...説明は後ほど...

- 生成的な記号や到達可能な記号の具体的な求めかた:
  のちほど 7.1.2 節で説明します

例7.1 (1/3)

文法 $G = (V, T, P, S)$:

- $S \ \rightarrow \ AB \mid a$
- $A \ \rightarrow \ b$

B は生成的でない

- a, b : 自分自身を生成(すると考える)
- S : a を生成
- A : b を生成

例7.1 (2/3)

変数集合VからBを除去

生成規則集合P から B を含む生成規則を除去

これで得られる文法 $G' = (V', T, P', S)$:

- $S \ \rightarrow \ a$
- $A \ \rightarrow \ b$

A は S から到達可能でない

例7.1 (3/3)

変数集合VからAを除去

生成規則集合P から A を含む生成規則を除去

これで得られる文法 $G'' = (V'', T, P'', S)$:

- $S \to a$

有用な記号の求め方の手順: 逆順にした場合

逆の手順

- 到達可能でない記号を除去
- 生成的でない記号を除去

逆手順でやってみた

文法 $G = (V, T, P, S)$:

- $S \to AB \mid a$
- $A \to b$

最初に到達可能でない記号を除去

- 全ての記号が到達可能なので記号は1つも除去されない

つぎに生成的でない記号を除去

- Bは生成的でないので除去

得られた文法

- $S \to a$
- $A \to b$ ... 要らないものが残ってる

定理7.2: 除去手順の正当性

文法 $G_1 = (V_1, T_1, P_1, S)$:
$G = (V, T, P, S)$ (ただし $L(G) \neq \emptyset$) に対して
以下の手順で得られる文法

1. Gより生成的でない記号(と関連する生成規則)を除去
   G_2 : これで得られた文法
2. G_2 より到達可能でない記号(と関連する生成規則)を除去
3. G_1 : 最終的に得られた文法

✔ $L(G) \neq \emptyset$ なので S は除去されない

定理 7.2
G_1には無用な記号がなく, $L(G_1) = L(G)$ が成立

定理7.2の証明: 概要

以下の2つを示せばよい

- G_1 に残った記号はいずれも有用
- $L(G_1) = L(G)$

定理7.2の証明: G_1 に残った記号はいずれも有用(1/3)

記号$X \in V_1 \cup T_1$: 除去されずG_1に残った任意の記号

X は第1ステップで除去されなかった
$=$ X は G の生成的な記号

- $\exists w \in T^* : X \overset{*}{\underset{G}{\Rightarrow}} w$
- この導出は G_2 の導出でもある: $X \overset{*}{\underset{G_2}{\Rightarrow}} w$
- この導出で現れる記号はすべてG_2で生成的なので

定理7.2の証明: G_1 に残った記号はいずれも有用(2/3)

X は第2ステップでも除去されなかった
　$= X$ は G_2 のS から到達可能な記号

- $S \overset{*}{\underset{G_2}{\Rightarrow}} \alpha X \beta$ となる α と β が存在
- この導出は G_1 の導出でもある: $S \overset{*}{\underset{G_1}{\Rightarrow}} \alpha X \beta$
- この導出で現れる記号はすべてG_2で到達可能なので

定理7.2の証明: G_1 に残った記号はいずれも有用(3/3)

$\alpha X \beta$ は終端記号列を生成

- $\alpha X \beta$に現れる各記号: G_1で到達可能
- しかも $V_2 \cup T_2$に属するのでG_2で生成的

導出$\alpha X \beta \overset{*}{\underset{G_2}{\Rightarrow}} xwy$ に含まれる記号は全て
G_2でSから到達可能

- この導出は G_1 の導出でもある

従って$S \overset{*}{\underset{G_1}{\Rightarrow}} \alpha X \beta \overset{*}{\underset{G_1}{\Rightarrow}} xwy$

結論: 各$X \in V_1 \cup T_1$ はG_1で有用

- G_1は無用な記号を持たない

定理7.2の証明: $L(G_1) = L(G)$

$L(G_1) \subseteq L(G)$ の証明

- G_1 は G から記号や生成規則を(0個以上)除去したもの
- 所属する語が増えはしない

$L(G_1) \supseteq L(G)$ の証明

- $w \in L(G)$ ならば$S \overset{*}{\underset{G}{\Rightarrow}} w$
- この導出に現れる記号:
  いずれも生成的かつ到達的なので G_1 に含まれる
- よって G_1 の導出でもある: $S \overset{*}{\underset{G_1}{\Rightarrow}} w$
  - ✔ 注意: 用いる生成規則はいずれも除去されていない
- 従って $w \in L(G_1)$

【証明おわり】

7.1.2 生成的記号と到達可能記号の計算

生成的記号の計算

文法 $G = (V, T, P, S)$ の生成的記号の計算方法

生成的記号の集合 $Z(\subseteq V \cup T)$ の計算手順

基礎:

- T の各要素を Z に加える
- 理由: T の各要素は自分自身を生成する(と考える)

帰納:

- 生成規則 $A \to \alpha$ においてαの全記号がZに属すとき:
  A を Z に加える
  - ✔ $A \to \epsilon$ の場合も含める
- Z に加えることのできる記号がなくなるまで繰り返す

例7.3

再掲: 例7.1の文法 $G = (V, T, P, S)$

- $S \rightarrow AB \mid a, \ A \rightarrow b$

基礎

- a, b は生成的
  - ✔ $Z = \{a, b\}$

帰納

- A は生成的: $A \rightarrow a$ において a は生成的なので
  - ✔ $Z = \{a, b, A\}$
- S は生成的: $S \rightarrow a$ において a は生成的なので
  - ✔ $Z = \{a, b, A, S\}$
- 条件に当てはまるものはもうない

結論: 生成的記号の集合: $Z = \{a, b, A, S\}$

定理7.4: 生成的記号の計算方法の正当性

X はアルゴリズムが求めた記号
$\Leftrightarrow$ Xは文法での生成的記号

証明略

到達可能記号の計算

文法 $G = (V, T, P, S)$ の到達可能記号の計算方法

到達可能記号の集合 $Z'(\subseteq V \cup T)$ の計算手順

基礎:

- S を Z' に加える
- 理由: S は開始記号なのでステップ0で到達可能

帰納:

- 記号AがZ'に属していれば
  生成規則 $A \rightarrow \alpha$ の右辺αの各記号をZ'に追加
- Z' に加えることのできる記号がなくなるまで繰り返す

例7.5

再掲: 例7.1の文法 $G = (V, T, P, S)$

- $S \rightarrow AB \mid a, \ A \rightarrow b$

基礎

- S は到達可能
  - ✔ $Z' = \{S\}$

帰納

- A, B, a は到達可能: $S \in Z'$ かつ $S \rightarrow AB|a$ より
  - ✔ $Z' = \{S, A, B, a\}$
- b は到達可能: $A \in Z'$ かつ $A \rightarrow b$ より
  - ✔ $Z' = \{S, A, B, a, b\}$
- 条件に当てはまるものはもうない

結論: 到達可能記号の集合: $Z' = \{S, A, B, a, b\}$

定理7.5: 到達可能記号の計算方法の正当性

X はアルゴリズムが求めた記号
$\Leftrightarrow$ Xは文法での到達可能記号

証明

- 定理7.4の証明と同様な手法でできる
- 各自で証明を行なってください

ε-規則の除去

ε-規則

ε-規則: $A \rightarrow \varepsilon$ の形の生成規則のこと

- 文法を作る際にあると便利

ε-規則は必須ではない

- 文脈自由言語 L はε-規則を含まない文法で生成可能
- ただし $\varepsilon \notin L$ の場合

これから示す事実:
　任意の文脈自由言語 L に対し
　$L - \{\varepsilon\}$ を生成する ε-規則なしの CFG が存在

消去可能な変数A

変数Aは消去可能とは: $A \stackrel{*}{\Rightarrow} \varepsilon$ のとき

消去可能な変数A, ε-規則の除去の考え方

観察例: 生成規則 $B \rightarrow CAD$
(ただし$A \stackrel{*}{\Rightarrow} \varepsilon$)

- A が ε を導出する可能性あり
- 導出しない可能性もあり
  ... 両方を考慮しないといけない

ε-規則の除去の方針

- 生成規則 $B \rightarrow CD$ を追加
  — Aから ε への導出の場合を除去
- 元の生成規則 $B \rightarrow CAD$ は残す
  — Aから ε 以外の導出の場合を残す
- 生成規則 $X \rightarrow \varepsilon$ の形のものは全て除去

文法 $G = (V, T, P, S)$ の消去可能変数の求め方

基礎: $A \rightarrow \varepsilon$ が G の生成規則ならば
　A は消去可能

帰納: $B \rightarrow C_1C_2 \cdots C_k$ が G の生成規則かつ
　$C_1, C_2, \cdots, C_k$ 全てが消去可能ならば
　B は消去可能

- 帰納段階では右辺が変数のみの生成規則を調べればよい

定理7.7: 消去可能変数の決定方法の正当性

任意の文法Gに対し先述の方法は
Gの消去可能変数のすべてを求める

証明略

例7.8: 文法からε-規則を除去 (1/3)

本例で用いる文法 G

- $S \to AB$
- $A \to aAA \mid \varepsilon$
- $B \to bBB \mid \varepsilon$

消去可能な変数を求める

- A と B : ε-規則を持つから
- S : 生成規則 $S \to AB$ にて A, B が消去可能なので

例7.8: 文法からε-規則を除去 (2/3)

G での生成規則 $S \to AB$ に注目

- A, B は消去可能
- $S \to AB \mid A \mid B$

G での生成規則 $A \to aAA$ に注目

- A は消去可能
- $A \to aAA \mid aA \mid a$

G での生成規則 $B \to bBB$ に注目

- B は消去可能
- $B \to bBB \mid bB \mid b$

例7.8: 文法からε-規則を除去 (3/3)

G での生成規則 $A \to \varepsilon$ に注目

- ε-規則なので除去

G での生成規則 $B \to \varepsilon$ に注目

- ε-規則なので除去

$X \to Y_1 Y_2 \cdots Y_k$ の形の生成規則はない

- ただし各 Y_i は消去可能

まとめ

- $S \to AB \mid A \mid B$
- $A \to aAA \mid aA \mid a$
- $B \to bBB \mid bB \mid b$

ε-規則の除去方法

文脈自由文法 $G = (V, T, P, S)$ から
ε-規則を除去した $G_1 = (V, T, P_1, S)$ の構成方法

1. G の消去可能変数を求める
2. G の生成規則 $A \to \varepsilon$:
   P_1 に追加しない
3. G の生成規則 $A \to \alpha$ (α は消去可能変数を含まない) :
   P_1 に追加
4. G の生成規則 $A \to \alpha$ ($|\alpha| \geq 1$) :
   右辺の消去可能変数を{含む/含まない}の組合せ
   全てに対する生成規則を P_1 に追加
   ただし $A \to \varepsilon$ は追加しない

定理7.9: ε-規則除去方法の正当性

先述のε-規則の除去方法に対し
$L(G_1) = L(G) - \{\varepsilon\}$ が成立

証明略(教科書参照のこと)

単位規則の除去

単位規則

単位規則: $A \to B$ (A, Bは変数)の形の生成規則のこと

- 文法を作る際にあると便利

単位規則は必須ではない

- 文脈自由言語 L は単位規則を含まない文法で生成可能

これら分かる事実2つ:

- 単位規則を<u>含む</u>CFGと等価な, 単位規則を<u>含まない</u>CFGが存在
- 任意の文脈自由言語 L に対し, Lを生成する単位規則を含まないCFGが存在

単位規則除去の基本的な考え方

$A \Rightarrow B_1 \Rightarrow B_2 \Rightarrow \cdots \Rightarrow B_k \Rightarrow \alpha$ のとき
生成規則 $A \to \alpha$ を追加

- 単位規則なしで A から直接 α を導出
- ただし各B_i は変数

用語定義: 単位対 (A, B)

- 単位規則だけを使って$A \stackrel{*}{\Rightarrow} B$ となるもの
- ただし A, B は変数

単位規則の除去の作業: 単位対の集合を元に行なう

単位対の集合の構成方法

基礎: (A, A) は単位対

- $A \stackrel{*}{\Rightarrow} A$ (長さ0の導出)なので

帰納: (A, B)が単位対かつ$B \to C$が単位規則なら (A, C) は単位対

- $A \stackrel{*}{\Rightarrow} B$ と規則 $B \to C$ より$A \stackrel{*}{\Rightarrow} C$

例7.10: 数式の文法(例5.27)の単位対 (1/2)

数式の文法(例5.27) 再掲

- $E \to T \mid E + T$
- $T \to F \mid T * F$
- $F \to I \mid (E)$
- $I \to a \mid b \mid Ia \mid Ib \mid I0 \mid I1$

基礎

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

帰納

- (E, T) は単位対: (E, E) と $E \to T$ より
  ...つづく...

例7.10: 数式の文法(例5.27)の単位対 (2/2)

帰納 (つづき)

- (E, F) は単位対: (E, T) と $T \to F$ より
- (E, I) は単位対: (E, F) と $F \to I$ より
- (T, F) は単位対: (T, T) と $T \to F$ より
- (T, I) は単位対: (T, F) と $F \to I$ より
- (F, I) は単位対: (F, F) と $F \to I$ より
- これ以上はもうない

定理7.11: 単位対の集合の構成方法の正当性

先述の方法は文脈自由文法Gの単位対を全て求める

証明略(教科書参照のこと)

単位規則の除去方法

文脈自由文法 $G = (V, T, P, S)$ から
単位規則を除去した $G_1 = (V, T, P_1, S)$ の構成方法

1. 単位規則 $A \rightarrow B$ は P_1 に追加しない
2. 各非単位規則 $A \rightarrow \alpha$ を P_1 に追加
3. G の単位対すべてを求める
4. 各単位対 (A, B) と非単位規則 $B \rightarrow \alpha$ に対し
   生成規則 $A \rightarrow \alpha$ を P_1 に追加

例7.12: 単位規則除去の例

教科書を見て理解しておいて下さい

定理7.13: 単位規則の除去方法の正当性

先述の単位規則の除去方法に対し
$L(G_1) = L(G)$ が成立

証明略(教科書参照のこと)

文法の単純化の方法のまとめ

単純化は以下の順番で行なう

- ε-規則の除去
- 単位規則を除去
- 無用な記号を除去

注意: 他の順序では完全に除去できない場合あり

定理7.14: 文法の単純化

G: ε 以外の列を少なくとも1つ生成する
任意の文脈自由文法

以下の条件を満たす文脈自由文法 G_1 が存在

- ε-規則を持たない
- 単位規則を持たない
- 無用な記号を持たない
- 同一の言語を生成: $L(G) = L(G_1)$

証明略

7.1.5 チョムスキーの標準形

チョムスキー標準形の定義

定義: チョムスキー標準形 (Chomsky Normal Form)
生成規則を以下の形に限定した文脈自由文法

- $A \to BC$ (ただし A, B, C は変数)
- $A \to a$ (ただし A は変数, a は終端記号)

チョムスキー標準形は文脈自由文法のサブクラス

- もし文法 G がチョムスキー標準形ならば
  G は文脈自由文法

※チョムスキー (人名):
文脈自由文法を考案した言語学者

定理7.16: CFGと等価なチョムスキー標準形の存在

文法G: 任意の文脈自由文法(CFG)で
ε 以外の列を少なくとも1つ生成するもの

定理: $L(G_1) = L(G) - \{\varepsilon\}$ を満たす
チョムスキー標準形文法 G_1 が存在

証明略(教科書参照のこと)

重要点: 「どんな文脈自由言語でも
チョムスキー標準形で文法が書ける」

チョムスキー標準形への変換方法 (1/4)

文法 G: 文脈自由文法

- 一般性を失うことなく
  ε-規則, 単位規則, 無用記号はないと仮定できる
- (もしある場合は前述の方法で除去できるため)

G の仮定より生成規則は以下のの形に限られる

1. 右辺は1つの終端記号だけ
   $A \to a \ (a \in T)$
2. 右辺は長さ2以上の記号(変数 and/or 終端記号)の列
   $A \to B_1B_2B_3 \cdots B_k \ (B_i \in V \cup T,\ k \geq 2)$

チョムスキー標準形への変換方法 (2/4): 前処理

右辺が長さ2以上の規則が終端記号が含む場合

- 右辺の中の各終端記号 a を対応する変数 A で置き換え
- 生成規則 $A \to a$ を追加

例: $X \to aXbY$ は以下のように変換

- $X \to AXBY$
- $A \to a$
- $B \to b$

この時点で得ている文法は以下の通り

- 右辺が長さ2以上の規則の右辺: 変数のみを含む
- 右辺が長さ1の規則の右辺: 終端記号
  - ✔ (仮定より) ε-規則, 単位規則, 無用記号はないから

チョムスキー標準形への変換方法 (3/4): 変換

(右辺の長さが1) 形 $A \to a$ の場合

- そのままチョムスキー標準形の生成規則として採用

(右辺の長さが2以上) 形 $A \to B_1B_2B_3 \cdots B_k$ の場合:
更に場合分け

- 右辺が長さ2の場合 ($k = 2$): そのまま規則として採用
- 右辺が長さ3以上の場合 ($k \geq 3$): つぎのスライドで説明

チョムスキー標準形への変換方法 (4/4): 右辺長が3以上

形 $A \to B_1B_2B_3 \cdots B_k \ (k \geq 3)$

- 新たに変数 $C_1, C_2, ...C_{k-2}$ を導入
- 以下の生成規則を採用
  $A \to B_1C_1$
  $C_1 \to B_2C_2$
  $C_2 \to B_3C_3$
  $\cdots$
  $C_{k-2} \to B_{k-1}B_k$

まとめ

- 右辺長が1の生成規則: 右辺には終端記号
- 右辺長が2の生成規則: 右辺には2つの変数
- 右辺長が3以上の生成規則: ない

チョムスキー標準形への変換方法の正当性

教科書を読んでおいて下さい

例7.15

教科書を読んでおいて下さい

おわり

ミニレポート (CLE)

教科書300ページ 問7.1.4