計算論 A / 2015年度

第7章 文脈自由言語の性質 (3/3)

文脈自由言語の閉包性

角川　裕次

もくじ

第7章

【教科書 p309〜】

＊＊＊ 本日の重要概念 ＊＊＊

閉包性
各種の判定問題の決定可能性

φ(・ω・) メモメモ

7.3 文脈自由言語の閉包性

閉包性

閉包性

言語: 集合 (要素は語)

クラス C: 言語クラス (CFL など)

$L \in C$ への演算結果が C に属すか否かを調べる

- 例 (二項演算 f) $\forall L_1, L_2 \in C : f(L_1, L_2) \in C$ か否か

7.3.1 代入

代入

代入: 各 $a \in \Sigma$ に対して言語 L_a を対応づけ

- Σ: 言語 L のアルファベット

代入を表す記号 s の定義

- $s(a) = L_a$ (各 $a \in \Sigma$ に対して)
- $s(a_1 \cdots a_n)$
  $= \{x_1 \cdots x_n | x_1 \in s(a_1), ..., x_n \in s(a_n)\}$
- $s(L) = \bigcup_{w \in L} s(w)$

例7.22 (1/2)

$s(0) = L_0 = \{a^n b^n \mid n \geq 1\}$

$s(1) = L_1 = \{aa,\ bb\}$

$s(01) = s(0)s(1) = L_0 L_1$
$= \{a^n b^n aa,\ a^n b^n bb \mid n \geq 1\}$

例7.22 (2/2)

$L = L(0^*)$

$s(L) = s(0)^* = L_0^*$
$= \{a^{n_1} b^{n_1} a^{n_2} b^{n_2} \cdots a^{n_k} b^{n_k}$
$\mid k \geq 0,\ n_1, n_2, ..., n_k \geq 1\}$

要素の具体例

- ε
- ab
- $abaabb$
- $aabbaaabbb$
- $abaabbabab$

定理7.23: 代入演算の閉包性

$s(L)$ は文脈自由言語である

- L : Σ 上の文脈自由言語
- s : Σ 上の代入
- $s(a) = L_a$: 文脈自由言語 (各 $a \in \Sigma$ に対して)

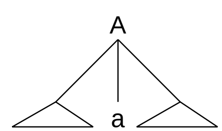
定理7.23: 証明の概要 (1/2)

Lの文脈自由文法を変更して $s(L)$ が生成できる

G: L を生成する文脈自由文法

各 $a \in \Sigma$ を生成するGの生成規則$A \rightarrow \cdots a \cdots$

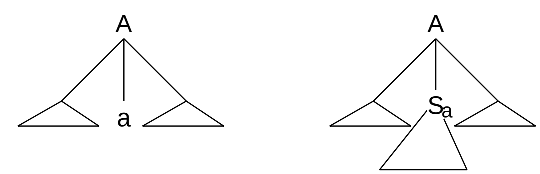
- 導出木の一部



定理7.23: 証明の概要 (2/2)

$A \rightarrow \cdots a \cdots$ を
$A \rightarrow \cdots S_a \cdots$ に変更

- S_a: $s(a)$ を生成する文脈自由文法の出発記号
- 元の文法とは, 変数等は互いに素と仮定



a の代わりに $s(a)$ の言語の語が導出される

【証明概要おわり】

7.3.2 代入定理の応用

定理7.24

文脈自由言語は次の演算のもとで閉じている

- 集合和
- 連接
- 閉包 $*$
- 正閉包 $+$
- 準同型写像

代入演算の閉包性の結果を用いると証明が簡単

- 代入演算の特殊の形なので証明を直ちに得る
- 演算ごとに別証明をしなくてすむ

定理7.24の証明: 集合和

L_1, L_2: 任意の文脈自由言語

$L_1 \cup L_2 = s(L_0)$ は文脈自由言語

- $s(1) = L_1$
- $s(2) = L_2$
- $L_0 = \{1, 2\}$ (... これは文脈自由言語)
- $s(L_0) = \{w_i\ ,\ w_j \mid w_i \in L_1, w_j \in L_2\} = L_1 \cup L_2$

定理7.24の証明: 連接

L_1, L_2: 任意の文脈自由言語

$L_1 \cdot L_2 = s(L_0)$ は文脈自由言語

- $s(1) = L_1$
- $s(2) = L_2$
- $L_0 = \{12\}$ (... これは文脈自由言語)
- $s(L_0) = \{w_i w_j \mid w_i \in L_1, w_j \in L_2\} = L_1 \cdot L_2$

定理7.24の証明: 閉包 $*$

L: 任意の文脈自由言語

$L^* = s(L_0)$ は文脈自由言語

- $s(1) = L$
- $L_0 = \{1^*\}$ (... これは文脈自由言語)
- $s(L_0) = \{w_i^* \mid w_i \in L\} = L^*$

定理7.24の証明: 正閉包 $+$

L: 任意の文脈自由言語

$L^+ = s(L_0)$ は文脈自由言語

- $s(1) = L$
- $L_0 = \{1^+\}$ (... これは文脈自由言語)
- $s(L_0) = \{w_i^+ \mid w_i \in L\} = L^+$

定理7.24の証明: 準同型写像

L: アルファベット Σ 上の任意の文脈自由言語
h: Σ 上の準同型写像

$h(L)$ は文脈自由言語

- 各 $a \in \Sigma$ に対し $s(a) = \{h(a)\}$ とすればよい
- このとき $h(L) = s(L)$

【証明おわり】

復習: 準同型写像 (教科書154ページ)

- 文字列上で定義される関数(写像)のこと
- 文字列上の各文字を特定の文字列で置き換える

7.3.3 逆順

定理7.25

L: 任意の文脈自由言語

L^R は文脈自由言語

生成規則の右辺を逆順にすればよい :
$A \to BC$ を $A \to CB$ にする

7.3.4 正則言語との共通部分

例7.26: 文脈自由言語は共通部分に関して閉じていない

L_1, L_2: 任意の文脈自由言語

$L_1 \cap L_2$ は文脈自由言語とは限らない

反例が存在

- $L_1 = \{0^n 1^n 2^i \mid n \geq 1, i \geq 1\}$: 文脈自由言語
- $L_2 = \{0^i 1^n 2^n \mid n \geq 1, i \geq 1\}$: 文脈自由言語
- $L_1 \cap L_2 = \{0^n 1^n 2^n \mid n \geq 1\}$ は文脈自由言語ではない
  (← 反復補題のところでやりました)

定理7.27: 文脈自由言語と正則言語との共通部分

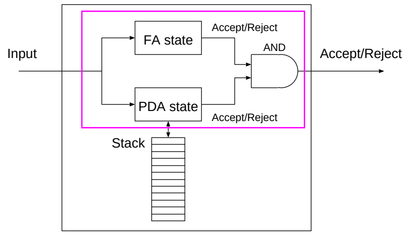
L: 任意の文脈自由言語
R: 任意の正則言語

$L \cap R$: は文脈自由言語である

定理27: 証明概略

L を受理するPDAの中に
R を受理するFAを組み込んで並行実行

- PDA の制御部 = 内部PDAの制御部 ＋ 内部FAの制御部



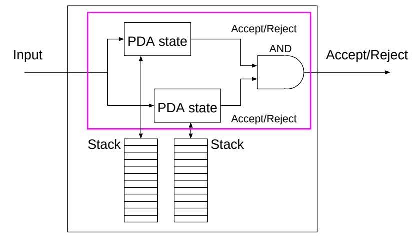
内部PDA と 内部FA がともに受理すると入力を受理

定理27: 検討

Q: PDAを2つ組み込めば共通部分を受理できるよね

A: PDA を2つ組み込んでみた:

- スタックが2つ: これは PDA ではない
- (2スタック = チューリングマシンと同等な計算能力)



定理7.29 (1)

L: 任意の文脈自由言語

R: 任意の正則言語

$L - R$ は文脈自由言語である

$L - R = L \cap \overline{R}$ の関係より成立

- R が正則言語ならば $\overline{R}$ も正則言語

定理7.29 (2)

L: 任意の文脈自由言語

$\overline{L}$ (Lの補集合)は文脈自由言語とは限らない

$L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$ の関係あり

- ドモルガンの法則

もし補集合演算が閉じていると仮定
　　共通集合演算も閉じていることになる
　　既知の結果に矛盾

定理7.29 (3)

L_1, L_2: 任意の文脈自由言語

$L_1 - L_2$ は文脈自由言語とは限らない

背理法で示す

- $L_1 - L_2$ が文脈自由言語だと仮定
- $L_1 = \Sigma^*$ とおく (L_1 は文脈自由言語)
- $L_1 - L_2 = \overline{L_2}$ なので $\overline{L_2}$ は文脈自由言語
- 矛盾 (補集合演算に関して閉じていないから)

7.3.5 逆準同型写像

復習: 逆準同型写像

言語Lの逆準同型写像 $h^{-1}(L) = \{w \mid h(w) \in L\}$

- h： 準同型写像

定理7.30

L: 任意の文脈自由言語

h: 準同型写像

$h^{-1}(L)$ は文脈自由言語である

定理7.30: 証明概要 (1/2)

仮定 (L は文脈自由言語) より
$h(a_1)h(a_2)\cdots h(a_n) \in L$ を受理するPDAが存在

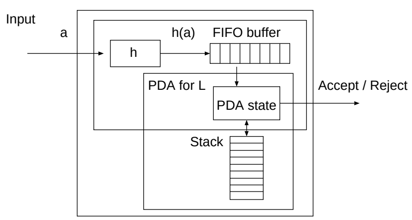
記号列 $a_1a_2\cdots a_n$ を受理する PDA を構成できる

- 内部でLを受理するPDAを動作させる

定理7.30: 証明概要 (2/2)

記号列 $a_1a_2\cdots a_n$ を受理する PDA の構成

- 空バッファ時は記号a_iを読み$h(a_i)$をバッファに挿入
- バッファから1文字読んでLを受理する内部PDAを動かす



7.4 文脈自由言語の決定問題

7.4.1 CFG と PDA の間の変換の複雑さ (結果だけを紹介)

入力長の表記

n: 入力の長さ

- 各種の変換アルゴリズム, 決定アルゴリズムへの入力
- PDA や CFG を文字列で記述した長さ

n の関数の形で実行時間を表記

定理 7.31

長さ n の記述の PDA P から,
高々長さ $O(n^2)$ の CFG G を生成する
$O(n^3)$ 時間アルゴリズムが存在

7.4.2 チョムスキー標準形への変換
(結果だけを紹介)

定理7.32

長さ n の文法 G に対し G と等価な
チョムスキー標準形は $O(n^2)$ 時間で求まる.
結果の文法は長さ $O(n^2)$.

7.4.3 文脈自由言語の空集合検査
(結果だけを紹介)

文脈自由言語の空集合検査問題

文法 G: 任意の文脈自由文法

S: G の出発記号

問題: S は列を少なくとも1つ生成するか否かを判定

文脈自由言語の空集合検査アルゴリズム

7.1.2項で検査アルゴリズムを示した
実行時間は $O(n^2)$ 時間 — 素朴な方法の場合

データ構造の工夫で $O(n)$ 時間で実現できる

- 教科書参照

7.4.4 CFL への所属検査

CFLへの所属検査問題

P: 任意に与えられたPDA

問題: 任意に与えられた記号列 w に対し
$w \in L$ か否かを判定

- ただし L は P の受理する文脈自由言語

CFLへの所属検査のための CKY アルゴリズム

$O(n^3)$ 時間で判定

- $n = |w|$, w は入力記号列
- PDA P の大きさは定数とみなす

仮定: チョムスキー標準形 G が与えられる
(PDAの代わりに)

- PDA から CFG は $O(n^3)$ 時間で構成可能
- CFG からチョムスキー標準形に $O(n^2)$ 時間で構成可能
- 従って上記仮定は一般性を失わない

FAQ: 「CKY」って何？
— 考案者3名 Cocke, Younger, 嵩 (元阪大基礎工教授) の名前より

CKY アルゴリズム (1/2)

入力 w を $w = a_1 a_2 \cdots a_n$ とおく

- 各 a_i は終端記号

表を作り判定

- 横軸の各 a_i は w の各文字 (例は $n = 5$ の場合)
- 表の要素 $X_{i,j}$ は変数 A の集合
  ただし $A \overset{*}{\Rightarrow} a_i a_{i+1} \cdots a_j$ となるものすべて

X15				
X14	X25			
X13	X24	X35		
X12	X23	X34	X45	
X11	X22	X33	X44	X55
a1	a2	a3	a4	a5

CKY アルゴリズム (2/2)

最終的な判定: $S \in X_{1,n} \Leftrightarrow w \in L$

- S: G の出発記号
- $S \in X_{1,n}$ は $S \overset{*}{\Rightarrow} a_1 a_2 \cdots a_n = w$ と同値
- つまり $w \in L$ か否かを判定できる

注意

- 表の要素 $X_{1,n}$ は $A \overset{*}{\Rightarrow} a_1 a_2 \cdots a_n$ となる変数 A の集合
- $S \in X_{1,n}$ なら $S \overset{*}{\Rightarrow} a_1 a_2 \cdots a_n$ だと判定できる

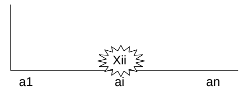
CKY アルゴリズム: 表の構成の概要

行毎に下から上に向かって埋めてゆく

CKY アルゴリズムの表の構成 (1/3)

基礎 (表の最初の行, すなわち各 $X_{i,i}$)

- 求めるべきは $A \stackrel{*}{\Rightarrow} a_i$ となる変数 A の集合
- A の可能性は生成規則 $A \to a_i$ で判定可能
- $X_{i,i}$ をそのような変数 A の集合とする

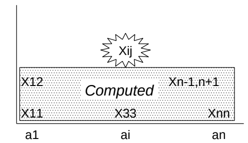


- ※文法はチョムスキー標準形を仮定

CKY アルゴリズムの表の構成 (2/3)

帰納 (各 $X_{i,j}$): 表の下の行はすべて計算済みと仮定

- $A \stackrel{*}{\Rightarrow} a_i a_{i+1} \cdots a_j$ となる変数 A 全てを求め $X_{i,j}$ とする



観察: 導出 $A \stackrel{*}{\Rightarrow} a_i a_{i+1} \cdots a_j$ は $A \Rightarrow BC$ から始まる

- ある k に対し $B \stackrel{*}{\Rightarrow} a_i a_{i+1} \cdots a_k$, $C \stackrel{*}{\Rightarrow} a_{k+1} a_{k+2} \cdots a_j$

表の計算済み部分から $X_{i,j}$ を構成 (次スライド)

CKY アルゴリズムの表の構成 (3/3) つづき

帰納 (つづき): $X_{i,j}$ の構成

$X_{i,j}$ は以下の条件を満たす変数 A の集合

1. G の生成規則 $A \to BC$ と $i \leq k < j$ それぞれに対して
2. $B \in X_{i,k}$ $(B \stackrel{*}{\Rightarrow} a_i a_{i+1} \cdots a_k$ の関係)
3. $C \in X_{k+1,j}$ $(C \stackrel{*}{\Rightarrow} a_{k+1} a_{k+2} \cdots a_j$ の関係)

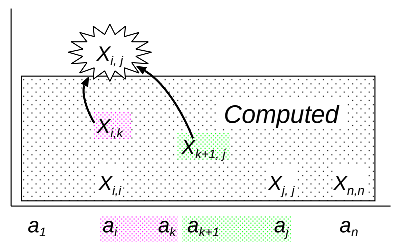
そのような A の見つけ方

- $(X_{i,i}, X_{i+1,j}), (X_{i,i+1}, X_{i+2,j}), \ldots, (X_{i,j-1}, X_{j,j})$ を調べれば良い
- これら X はいずれも計算済み

CKY アルゴリズムの表の構成 (3'/3) 図説

$A \in X_{i,j}$ とするのは以下の時

- $A \to BC$ が生成規則
- ある k に対し $B \in X_{i,k}$
- ある k に対し $C \in X_{k+1,j}$



CKY アルゴリズム: まとめ

定理7.33 CKY アルゴリズムは $O(n^3)$ 時間で所属検査ができる

注1: CKYアルゴリズムは任意の文脈自由言語が対象

注2: コンパイラでの実際

- 文脈自由言語の部分クラスを対象
- 高速に構文解析を行なう

7.4.5 決定不能な CFL 問題のあらまし
(結果だけを紹介)

決定可能性と決定不能性

決定可能な問題

- 有限時間で答(Yes/No)を出力して
停止するアルゴリズムが存在する問題

決定不能な問題

- 決定可能でない問題, すなわち
- 判定に要する時間が有限時間で行なえない問題

CFLに関する決定不能な問題

与えられた CFG はあいまいか

与えられた CFL は本質的にあいまいか

2つの CFL の共通部分は空集合か

2つの CFL は等しいか

与えられた CFL が Σ^* に等しいか

おわり

ミニレポート (CLE)

教科書332ページ 問7.4.4