

アルゴリズムとデータ構造III 2回目: 10月09日

チューリング機械, 文脈自由文法

授業資料 <http://ir.cs.yamanashi.ac.jp/~ysuzuki/algorithm3/index.html>

授業の予定(中間試験まで)

1	10/02	スタック(後置記法で書かれた式の計算)
2	10/09	チューリング機械, 文脈自由文法
3	10/16	構文解析 CYK法
4	10/23	構文解析 CYK法
5	10/30	構文解析 CYK法
6	11/06	構文解析 チャート法
7	11/13	グラフ(動的計画法, ダイクストラ法, DPマッチング)
8	11/20	グラフ(DPマッチング, ビームサーチ, A*アルゴリズム)
9	11/27	中間試験

授業の予定(中間試験以降)

10	12/04	全文検索アルゴリズム (simple search, KMP)
11	12/11	全文検索アルゴリズム (BM, Aho-Corasick)
12	12/18	全文検索アルゴリズム (Aho-Corasick), データ圧縮
13	01/08	暗号(黄金虫, 踊る人形) 符号化(モールス信号, Zipfの法則, ハフマン符号)テキスト圧縮
14	01/15	テキスト圧縮 (zip), 音声圧縮 (ADPCM, MP3, CELP), 画像圧縮 (JPEG)
15	01/29	期末試験

形式言語と有限オートマトン入門 4.5.2 チューリング機械 重要!

- 言語受理能力が最も高いオートマトン
- 半無限長の読み書きが自由に行えるテープを用いた有限状態機械

読み書きテープ (初期状態では入力語が記述されている)

0	1	1	0	0	1	B	B	B	B	...
---	---	---	---	---	---	---	---	---	---	-----

読み書きヘッド (初期状態: 左端 語の先頭文字位置)
テープ上を左右に移動, read, rewrite

有限状態制御部
最終状態に遷移すると停止して入力語を受理する

チューリング機械(TM)の定義

TM $M=(Q, \Sigma, \Gamma, \delta, S, B, F)$

- Q : 内部状態の集合
- Σ : 入力アルファベット B を含まない
- Γ : テープ記号の集合 ($\Gamma \supset \Sigma$)
- B : 空白記号 Γ の要素であるが Σ の要素ではない
- δ : 状態遷移関数 $\delta: Q \times \Gamma \rightarrow Q \times \Gamma \times \{R, S, L\}$
- R : ヘッドを右に移動, S : ヘッドを移動させない, L : ヘッドを左に移動
- S : 初期状態 $S \in Q$
- F : 最終状態(受理状態)の集合 $F \subset Q$

例題4.71 $w_1=0101$

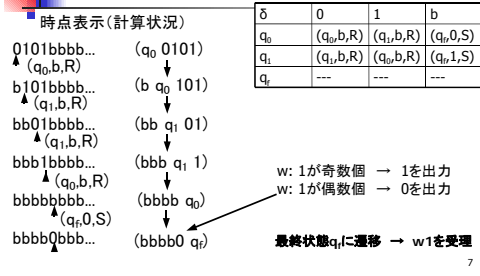
$Q=\{q_0, q_1, q_f\}, \Sigma=\{0,1\}, \Gamma=\{0,1,b\}, S=q_0, B=b, F=\{q_f\}$

δ	0	1	b
q_0	(q_0, b, R)	(q_1, b, R)	$(q_f, 0, S)$
q_1	(q_1, b, R)	(q_0, b, R)	$(q_f, 1, S)$
q_f	---	---	---

計算状況を示せ。
 Σ^* 上の任意の語と、その最終計算状況におけるテープ上の記号との対応を答えよ

例題4.71 答え

w1=0101



7

例題4.71 w2'=011010

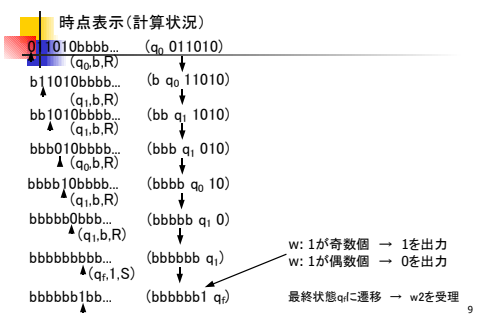
$Q=\{q_0, q_1, q_f\}, \Sigma=\{0, 1\}, \Gamma=\{0, 1, b\}, S=q_0, B=b, F=\{q_f\}$

δ	0	1	b
q_0	(q_0, b, R)	(q_1, b, R)	$(q_f, 0, S)$
q_1	(q_1, b, R)	(q_0, b, R)	$(q_f, 1, S)$
q_f	---	---	---

計算状況を示せ。
 Σ^* 上の任意の語と、その最終計算状況におけるテープ上の記号との対応を答えよ

8

例題4.71 答え W2'=011010



9

練習問題1

例題4.71 w2=01101

$Q=\{q_0, q_1, q_f\}, \Sigma=\{0, 1\}, \Gamma=\{0, 1, b\}, S=q_0, B=b, F=\{q_f\}$

δ	0	1	b
q_0	(q_0, b, R)	(q_1, b, R)	$(q_f, 0, S)$
q_1	(q_1, b, R)	(q_0, b, R)	$(q_f, 1, S)$
q_f	---	---	---

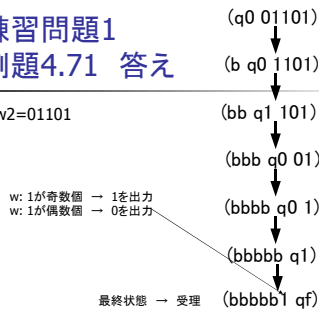
計算状況を示せ。
 Σ^* 上の任意の語と、その最終計算状況におけるテープ上の記号との対応を答えよ

10

練習問題1

例題4.71 答え

w2=01101



11

4.5.3 オートマトンと計算理論

オートマトンの受理する言語クラス

オートマトン	受理言語型	言語クラス
チューリング機械	第0型言語	句構造言語(PSL)
線形拘束チューリング機械	第1型言語	文脈依存言語(CSL)
プッシュダウンオートマトン	第2型言語	文脈自由言語(CFL)
有限オートマトン	第3型言語	正規言語(RL)

$RL \subset CFL \subset CSL \subset PSL$ (チョムスキーの言語階層)¹²

万能チューリングマシン

- 任意のTMIについて、その動作表を与えられるとあたかもそのTMのように振る舞うTM
- コンピュータ
 - プログラム=動作表(状態遷移関数表)
 - 入力=入力語
 - コンピュータは万能TM
- チューリングテスト
 - TM M が人間
 - コンピュータ(TM)がTM M を完全に模倣できるか

13

「形式言語と有限オートマトン入門」 5 形式言語理論入門

- 5.1 形式言語理論
- 5.2 文脈自由文法
- 5.3 線形文法と正規言語
- 5.4 形式言語のクラス階層とオートマトン
- 5.5 言語処理への応用

14

形式文法Gの定義

- $G=(N,T,P,S)$
 - N: 非終端記号の集合
 - T: 終端記号の集合
 - P: プロダクション
 - S: 開始記号

15

5.2 文脈自由文法

- 文脈自由文法(CFG)
 - 文脈自由プロダクションのみから構成される
 - 文脈自由プロダクション
 - $\alpha \rightarrow \beta$
 - ただし、 $\alpha \in N, \beta \in V^*$
 - N: 非終端記号の集合, T: 終端記号の集合, V: NとTの直和
 - 左辺が変数1つ
- 文脈依存文法(CSG)
 - 文脈依存プロダクションを含むプロダクションから構成される
 - 文脈依存プロダクション
 - $u\alpha v \rightarrow u\beta v$ ただし、 $\alpha \in N, u, v \in V^*, \beta \in V^*$
 - N: 非終端記号の集合, T: 終端記号の集合, V: NとTの直和
 - $u=v=\epsilon$ のとき $(\alpha \rightarrow \beta)$ 文脈自由プロダクションとなる

16

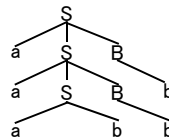
文脈自由文法の例(例題5.9)

- CFG $G=(N,T,P,S)$
 - N(非終端記号) = $\{B, S\}$
 - T(終端記号) = $\{a, b\}$
 - P: $S \rightarrow aSB \mid ab$
 $B \rightarrow b$
- 語 $aaabbbb$ の導出過程
- $L(G)$ はどのような言語か

17

例題5.9の解答例

- CFG $G=(N,T,P,S)$
 - N = $\{B, S\}$
 - T = $\{a, b\}$
 - P: $S \rightarrow aSB \mid ab, B \rightarrow b$
- $S \rightarrow aSB \Rightarrow aaSB \Rightarrow aaabBB \Rightarrow aaabbbB \Rightarrow aaabbbb$
- $L(G): a^n b^n$
- 正規表現では表せない
- プッシュダウンオートマトンでは表現可能
- 構文木



18

練習問題2

例題5.10 文脈依存文法の例

- CSG $G=(N,T,P,S)$
- $N=\{A,B,S\}$
- $T=\{a,b\}$
- $P: S \rightarrow aSBA \mid abA, AB \rightarrow BA, bB \rightarrow bb, bA \rightarrow ba, aA \rightarrow aa$
- 語 aabbaa の導出過程
- $L(G)$ はどのような言語か

19

練習問題2 解答

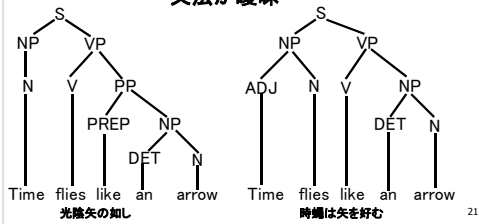
例題5.10 aabbaa

- CSG $G=(N,T,P,S)$
- $N=\{A,B,S\}$
- $T=\{a,b\}$
- $P: S \rightarrow aSBA \mid abA, AB \rightarrow BA, bB \rightarrow bb, bA \rightarrow ba, aA \rightarrow aa$
- 語 aabbaa の導出過程
- $S \rightarrow aSBA \rightarrow aabABA \rightarrow aabBAA \rightarrow aabbAA \rightarrow aabbaa$
- $L(G)$ はどのような言語か
- $L(G): a^n b^n a^n$

構文木(導出木)

- Time flies like an arrow.

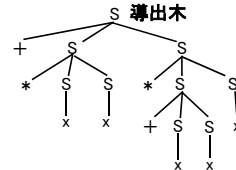
2種類の導出木
→ 文法が曖昧



21

例題5.11

- 問題:
- 文法 $N=\{S\}, T=\{x, +, *\}, P=\{S \rightarrow +SS \mid *SS \mid x\}$
- 語 $W = +*xx*+xxx$ を導出せよ
- 語 W の導出木

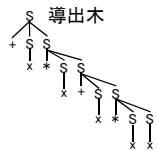


- 解答例
- 導出:
 $S \rightarrow +SS \rightarrow +*SSS \rightarrow +*xSS \rightarrow +*xxS \rightarrow +*xx*SS \rightarrow +*xx*+S$
 $SS \rightarrow +*xx*+xxx$

22

例題5.12 ①

- 問題
- 文法 $N=\{S\}, T=\{x, +, *\}, P=\{S \rightarrow +SS \mid *SS \mid x\}$
- 中置記法 $x+x*(x+x*x)$



- 解答例
- 前置記法 $+x*x+x*x*x$
- $S \rightarrow +SS \rightarrow +xS \rightarrow +x*SS \rightarrow +x*xS \rightarrow +x*x+SS$
 $\rightarrow +x*x+xS \rightarrow +x*x+x*xSS \rightarrow +x*x+x*x*xS \rightarrow +x*x+x*x*x*x$

23

練習問題3 例題5.12 ②

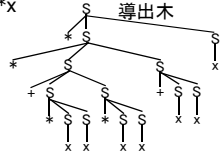
- 問題
- 文法 $N=\{S\}, T=\{x, +, *\}, P=\{S \rightarrow +SS \mid *SS \mid x\}$
- 中置記法 $(x*x+x*x)*(x+x)*x$

- 前置記法
- 最左導出
- 構文木

24

練習問題3 例題5.12 ②の解答例

- 問題
- 文法 $N = \{S\}, T = \{x, +, *\}, P = \{S \rightarrow +SS \mid *SS \mid x\}$
- 中置記法 $(x*x+x*x)*(x+x)*x$



- 解答例
- 前置記法 $**+*xx*xx+xxx$
- $S \rightarrow *SS \Rightarrow **SSS \Rightarrow **+SSSS \Rightarrow **+*SSSSS \Rightarrow **+*xSSSS$
- $\Rightarrow **+*xxSSS \Rightarrow **+*xx*SSSS \Rightarrow **+*xx*xSSS$
- $\Rightarrow **+*xx*xxSS \Rightarrow **+*xx*xx+SSS$
- $\Rightarrow **+*xx*xx+xxx$

25

文脈自由文法の曖昧性

- どのような導出を行っても同じ導出木が得られる
- \Rightarrow 文法Gは曖昧でない
- 複数の異なる導出木が構成できるような語を含む
- \Rightarrow 文法Gは曖昧である

26

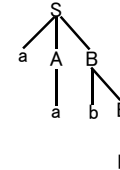
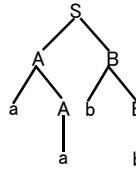
例題5.26

- 文法 $G = (N, T, P, S)$ において、 $N = \{S, A, B\}, T = \{a, b\}$,
- $P: S \rightarrow AB \mid aAB, A \rightarrow aA \mid a, B \rightarrow bB \mid b$
- この文法が曖昧であることを示せ

27

例題5.26 解答例 同一文字列に対して2種類の導出木が構成可能 \rightarrow 曖昧である

- 1. $S \rightarrow AB \rightarrow aAB \rightarrow aAbB \rightarrow aabB \rightarrow aabb$
- 2. $S \rightarrow aAB \rightarrow aaB \rightarrow aabB \rightarrow aabb$



28

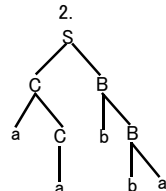
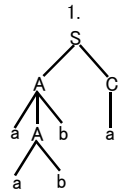
練習問題4 例題5.27

- 文法 $G = (N, T, P, S)$ において、
- $N = \{S, A, B, C\}, T = \{a, b\}$,
- $P: S \rightarrow AC \mid CB, A \rightarrow aA \mid a, B \rightarrow bB \mid ba$
- $C \rightarrow aC \mid a$
- この文法が曖昧であることを、 $aabba$ の導出木を構成して示せ

29

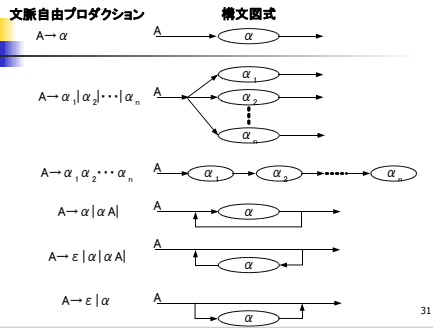
練習問題4 例題5.27 解答例 同一文字列に対して2種類の導出木が構成可能 \rightarrow 曖昧である

- 1. $S \rightarrow AC \rightarrow aAbC \rightarrow aAba \rightarrow aabba$
- 2. $S \rightarrow CB \rightarrow aCB \rightarrow aCbB \rightarrow aabB \rightarrow aabba$



30

CFGの構文図式



文脈自由文法の簡単化

- 以下の書き換え規則を削除する
- 詳しくは154ページ
 - 開始記号Sから導出に使われることのない非終端記号
 - ϵ -規則 ($A \rightarrow \epsilon: A \in N$)
 - 単位規則 ($A \rightarrow B: A, B \in N$)

32

例題5.31 ① ϵ -規則の消去

- $S \rightarrow aA, A \rightarrow bA | \epsilon$
 - $S \rightarrow aA$
 - $A \rightarrow bA$
 - $A \rightarrow \epsilon$
- $S \rightarrow aA | a$
- $A \rightarrow bA | b$
- $S \rightarrow aA | a, A \rightarrow bA | b$

33

例題5.31 ② ϵ -規則の消去

- $S \rightarrow aA, A \rightarrow aA | bB, B \rightarrow bB | \epsilon$
 - $S \rightarrow aA$
 - $A \rightarrow aA$
 - $A \rightarrow bB$
 - $B \rightarrow bB$
 - $B \rightarrow \epsilon$
- $S \rightarrow aA$
- $A \rightarrow aA$
- $A \rightarrow bB | b$
- $B \rightarrow bB | b$
- $S \rightarrow aA, A \rightarrow aA | bB | b, B \rightarrow bB | b$

34

例題5.31 ③ ϵ -規則の消去

- $S \rightarrow aAB, A \rightarrow aA | a | bB, B \rightarrow bB | \epsilon$
 - $S \rightarrow aAB$
 - $A \rightarrow aA$
 - $A \rightarrow a$
 - $A \rightarrow bB$
 - $B \rightarrow bB$
 - $B \rightarrow \epsilon$
- $S \rightarrow aAB$
- $A \rightarrow aA$
- $A \rightarrow a$
- $A \rightarrow bB | b$
- $B \rightarrow bB | b$
- $S \rightarrow aAB, A \rightarrow aA | a | bB | b, B \rightarrow bB | b$

35

例題5.32 ①

- $S \rightarrow aA, A \rightarrow aB | B, B \rightarrow bB | b$

36

例題5.33 ①

- $S \rightarrow AB|a, A \rightarrow a$
 - $S \rightarrow AB$
 - $S \rightarrow a$
 - $A \rightarrow a$
- $B \rightarrow$ が無いので $S \rightarrow AB$ を削除
- $S \rightarrow a$

37

文脈自由文法の標準形

- チョムスキー標準形
 - 文脈自由文法の規約化された生成規則が,
 - すべて $A, B, C \in N, a \in T$ として,
 - $A \rightarrow BC$ または $A \rightarrow a$ の形をしているとき, この生成規則をチョムスキー標準形という

38

文脈自由な生成規則のチョムスキー標準形への変換

- $X, A, B, C \in N, a \in T$ として,
- $X \rightarrow aB$ ならば $X \rightarrow AB, A \rightarrow a$ と分解する
- $X \rightarrow ABC$ ならば $X \rightarrow AY, Y \rightarrow BC$ と分解する

39

例題5.34

- 文法 $G = (N, T, P, S)$ において, $N = \{S, A, B, C\}, T = \{a, b\}, P$ を, $S \rightarrow AaC|CbBa, A \rightarrow aAb|ab, B \rightarrow bB|b, C \rightarrow Ca|a$ とする. この文法 G を
 - チョムスキー形生成規則をもつ文脈自由文法に書き換えよ.

40

例題5.34 解答例

- $S \rightarrow AaC \Rightarrow S \rightarrow AS_1, S_1 \rightarrow S_2C, S_2 \rightarrow a$
- $S \rightarrow CbBa \Rightarrow S \rightarrow CS_3, S_3 \rightarrow S_4S_5, S_4 \rightarrow b, S_5 \rightarrow BS_2$
- $A \rightarrow aAb \Rightarrow A \rightarrow S_2A_1, A_1 \rightarrow AS_4$
- $A \rightarrow ab \Rightarrow A \rightarrow S_2S_4$
- $B \rightarrow bB \Rightarrow B \rightarrow S_4B$
- $C \rightarrow Ca \Rightarrow C \rightarrow CS_2$

41

次回の講義 構文解析アルゴリズム

- CYK (Cocke-Younger-Kasami) 法
 - チョムスキー標準形で書かれた言語の構文解析手法
- チャート法, LR法

42