計算論 A / 2015年度

第7章 文脈自由言語の性質 (2/3)

文脈自由言語の反復補題

角川　裕次

第7章

＊＊＊ 本日の重要概念 ＊＊＊

文脈自由言語の反復補題

φ(・ω・)メモメモ

7.2 文脈自由言語の反復補題

どの文脈自由言語Lでも必ず有する性質 $P(L)$ を表明

任意の言語 L に対して

L が文脈自由言語ならば$P(L)$は真

— 対偶 —

$P(L)$が偽ならばLは文脈自由言語でない

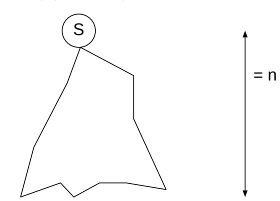
言語が文脈自由言語ではないことを示すための手法

7.2.1 構文木の大きさ

$G = (V, T, P, S)$: チョムスキー標準形の文法

$w \in L(G)$

n: wの構文木の最長路の長さ



定理: $|w| \leq 2^{n-1}$ が成立

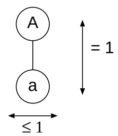
証明は n に関する帰納法

証明: 構文木の大きさ (1/2): 基礎 $n = 1$

$n = 1$: 根の子が葉だけの場合に相当

- 生成規則はチョムスキー標準形なので $A \to a$ の形
- 葉の数は1つだけ: $|w| = 1$

不等式成立: $|w| = 1 \leq 2^{n-1} = 2^0 = 1$

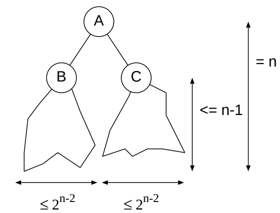


証明: 構文木の大きさ (2/2): 帰納 $n > 1$

構文木の根が使う生成規則は $A \to BC$ の形

- Bを根とする部分木: 最長路の長さは高々$n-1$
  帰納法の仮定より葉の数は高々2^{n-2}
- Cを根とする部分木: Bの場合と同様

Aを根とする部分木の葉の数: 高々 $2^{n-2} + 2^{n-2} = 2^{n-1}$



【証明おわり】

7.2.2 反復補題の表現

定理7.18: 文脈自由言語の反復補題

L: 任意の文脈自由言語

定数n: L によって決まる定数

z: Lに属する任意の列で $|z| \geq n$ であるもの

以下の条件を満たす z の分解 $z = uvwxy$ が存在

1. $|vwx| \leq n$
2. $vx \neq \varepsilon$　　(すなわち v と x は同時には ε でない)
3. 任意の$i (\geq 0)$に対し $uv^iwx^iy \in L$ が成立

証明の準備

G: Lを生成するチョムスキー標準形の文法

m: Gの変数の数

$n = 2^m$ と選ぶ

z: Lに属する任意の列で $|z| \geq n$ であるもの

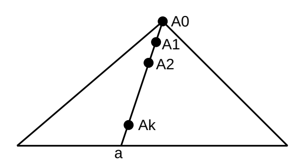
証明: 観察(1/3)

z に対する構文木: 長さ$m+1$以上の経路が存在

- 定理7.17: もし存在しないと $|z| \leq 2^{m-1}$ となり矛盾

長さ$m+1$以上の経路を任意に選ぶ

- 経路長さ: $k+1$ $(\geq m+1)$

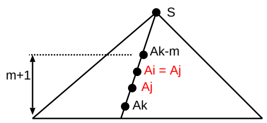


- 変数 $A_0, A_1, ..., A_k$: 選んだ経路に出現する変数
  (根から出現する順序で; 同一変数の重複出現を許す)

証明: 観察(2/3)

A_i と A_j の選び方

- 選んだ経路にともに出現
- 葉から根に向かう $m+1$ 個の変数に含まれる
- $A_i = A_j$ (同じ変数が異なる場所に重複して出現)



必ずそのような A_i, A_j が存在 (←鳩の巣原理による)

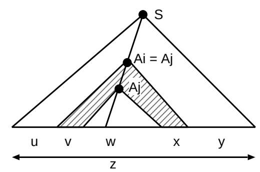
- m変数を重複を許して $m+1$ 箇所に割り当て
- 同じ変数が2回割り当てられるはず

証明: 観察(3/3)

図からわかること

- $S \stackrel{*}{\Rightarrow} uA_iy \stackrel{*}{\Rightarrow} uvwxy(=z)$
- $A_i \stackrel{*}{\Rightarrow} vA_jx \stackrel{*}{\Rightarrow} vwx$
- $A_j \stackrel{*}{\Rightarrow} w$

✔ u, v, w, x, y はいずれも終端記号の列

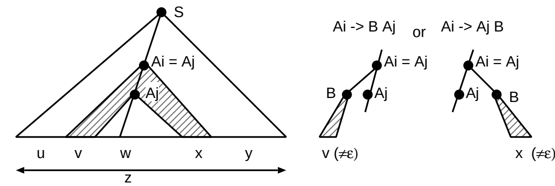


証明: 条件2

v と x はともに ε ではない

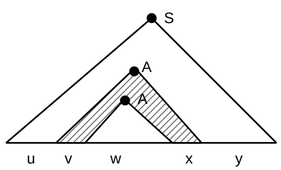
- 単位規則がないから
- ただし v と x のどちらか一方が ε の場合あり

従って $vx \neq \varepsilon$



証明: 条件3 (1/6)

$A = A_i = A_j$ とおく

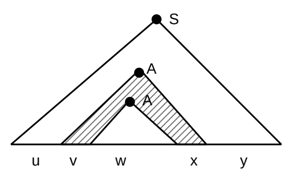


$uvwxy$ の導出:

- $S \stackrel{*}{\Rightarrow} u\underline{A}y \stackrel{*}{\Rightarrow} u\underline{vAx}y \stackrel{*}{\Rightarrow} uvwxy(=z)$

証明: 条件3 (2/6)

重要ポイント: $A \stackrel{*}{\Rightarrow} vAx$ に注目

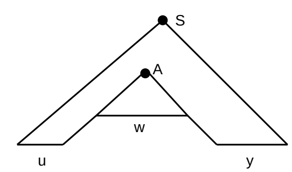


<u>!重要!：</u> $A \stackrel{*}{\Rightarrow} vAx$ を何度も繰り返した導出が可能

証明: 条件3, 可能な導出(3/6)

$S \stackrel{*}{\Rightarrow} uAy \stackrel{*}{\Rightarrow} uwy \;\; (\in L)$

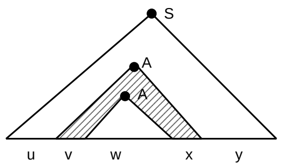
- 繰返し回数: 0 回



証明: 条件3, 可能な導出(4/6)

$S \stackrel{*}{\Rightarrow} uAy \stackrel{*}{\Rightarrow} uvAxy \stackrel{*}{\Rightarrow} uvwxy \ \ (= z \in L)$

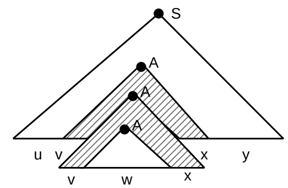
- ❖ 繰返し回数: 1 回



証明: 条件3, 可能な導出(5/6)

$S \stackrel{*}{\Rightarrow} uAy \stackrel{*}{\Rightarrow} uvAxy \stackrel{*}{\Rightarrow} uvvAxxy \stackrel{*}{\Rightarrow} uvvwxxy \ (\in L)$

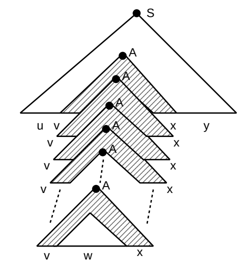
- ❖ 繰返し回数: 2 回



証明: 条件3, 可能な導出(6/6)

$S \stackrel{*}{\Rightarrow} uAy \stackrel{*}{\Rightarrow} uv^iAx^iy \stackrel{*}{\Rightarrow} uv^iwx^iy \ \ (\in L)$

- ❖ 繰返し回数: i 回



証明: 条件1

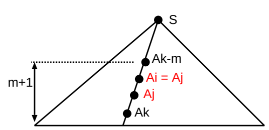
条件1: $|vwx| \le n$

$k - m \le i$ が成立

A_i を根とする部分木の最長路は長さ $m+1$ 以下

A_i から導出される vwx の長さは高々 $2^{(m+1)-1} = n$

- ❖ 定理7.17より



【証明おわり】

7.2.3 反復補題のCFLへの応用

例7.19: $L = \{0^i1^i2^i \mid i \ge 0\}$ (0/6)

注意: n を i に入れ換え: 教科書では

$$L = \{0^n1^n2^n \mid n \ge 0\}$$

と書いていますがスライドでは

$$L = \{0^i1^i2^i \mid i \ge 0\}$$

と書きます

- ❖ 反復補題で与えられる定数 n との混同を避けるため

例7.19: $L = \{0^i1^i2^i \mid i \geq 0\}$ (1/6)

L: 0, 1, 2 がともに同数

L は CFL ではない

証明は背理法: CFLだと仮定して矛盾を導く

n: 反復補題で与えられる定数

$z = 0^n1^n2^n$ を選ぶ

n n n

000 ... 000111 ... 111222 ... 222

例7.19: $L = \{0^i1^i2^i \mid i \geq 0\}$ (2/6)

z を $uvwxy$ に分割
(ただし $|vwx| \leq n$ かつ $vx \neq \varepsilon$)

観察: z の中の最後の0と最初の2は$n+1$離れている

n n n

000 ... 000111 ... 111222 ... 222

n+1

従って vwx が 0 と 2 をともに含むことはない

- $|vwx| \leq n$ なのでどうvwxを選んでもそうなる

例7.19: $L = \{0^i1^i2^i \mid i \geq 0\}$ (3/6)

Lに含むべきでない語がLに含まれることを示す

- 以下では場合分け

例7.19: $L = \{0^i1^i2^i \mid i \geq 0\}$ (4/6)

場合1: vwx が 2 を含まない場合

- 分割方法より $vx \neq \varepsilon$ なので $|vwx| > 0$

000 ... 000111 ... 111222 ... 222

v w x

uvwxy

0回の反復 $uv^0wx^0y = uwy$ を観察

- 0か1はn個未満
- 2はちょうどn個

uwyはLに含れない語
(反復補題を満たす分割でない)

例7.19: $L = \{0^i1^i2^i \mid i \geq 0\}$ (5/6)

場合2: vwx が 0 を含まない場合

- 分割方法より $vx \neq \varepsilon$ なので $|vwx| > 0$

000 ... 000111 ... 111222 ... 222

v w x

uvwxy

0回の反復 $uv^0wx^0y = uwy$ を観察

- 1か2はn個未満
- 0はちょうどn個

uwyはLに含れない語
(反復補題を満たす分割でない)

例7.19: $L = \{0^i1^i2^i \mid i \geq 0\}$ (6/6)

$z \in L$ のいかなる分割に対しても
繰返しで得られる列は L に含まれない

L は反復補題を満たさない

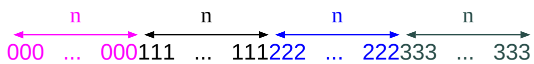
結論: L は文脈自由言語(CFL)ではない 【証明おわり】

例7.20: $L = \{0^i 1^j 2^i 3^j \mid i \geq 0, j \geq 0\}$ (1/5)

L: 0 と 2 が同数でしかも 1 と 3 が同数

n: 反復補題で与えられる定数

$z = 0^n 1^n 2^n 3^n$ を選ぶ

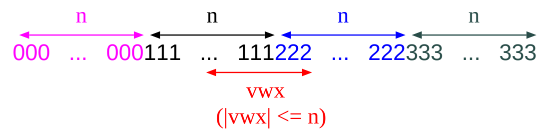


例7.20: $L = \{0^i 1^j 2^i 3^j \mid i \geq 0, j \geq 0\}$ (2/5)

z を $uvwxy$ に分割
(ただし $|vwx| \leq n$ かつ $vx \neq \varepsilon$)

観察1: vwx は1種類の記号だけ,
または隣接する2種類の記号で構成

- ❖ $|vwx| \leq n$ なのでどう vwx を選んでもそうなる



以下では2通りに分けて調べる: 出現文字が1種類 / 2種類

例7.20: $L = \{0^i 1^j 2^i 3^j \mid i \geq 0, j \geq 0\}$ (3/5)

場合1: vwx が1種類の記号だけで構成

0回の繰返し $uv^0wx^0y = uwy$ を観察

- ❖ 1種類の記号(vwxが含むもの)はn回未満の出現
- ❖ 他の3種類の記号はちょうどn回の出現

uwyはLに含まれない語

例7.20: $L = \{0^i 1^j 2^i 3^j \mid i \geq 0, j \geq 0\}$ (4/5)

場合2: vwx が2種類の隣接する記号で構成

0回の反復 $uv^0wx^0y = uwy$ を観察

場合2-1: vwxが 0 と 1 で構成される場合

- ❖ 0または1はn回未満の出現
- ❖ 2と3はn回の出現
  → uwyはLに含まれない語

場合2-2: vwxが 1 と 2 で構成される場合 — 同様

場合2-3: vwxが 2 と 3 で構成される場合 — 同様

例7.20: $L = \{0^i 1^j 2^i 3^j \mid i \geq 0, j \geq 0\}$ (5/5)

$z \in L$ のいかなる分割に対しても
繰返しで得られる列は L に含まれない

L は反復補題を満たさない

結論: L は文脈自由言語(CFL)ではない 【証明おわり】

例7.21: $L = \{ss \mid s \in \{0,1\}^*\}$ (0/6)

注意: n を i に入れ換え：教科書では

$$L = \{ww \mid w \in \{0,1\}^*\}$$

と書いてますがスライドでは

$$L = \{ss \mid s \in \{0,1\}^*\}$$

と書きます

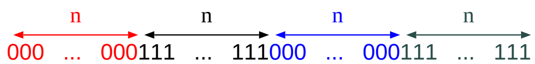
- ❖ 反復補題での分割 $z = uvwxy$ での w との
  混同を避けるため

例7.21: $L = \{ss \mid s \in \{0,1\}^*\}$ (1/6)

L: 2つの同一の記号列が並んだ言語

n: 反復補題で与えられる定数

$z = 0^n 1^n 0^n 1^n$ を選ぶ

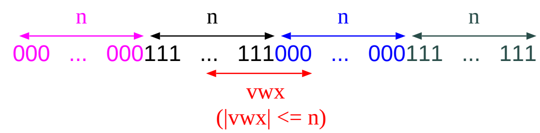


例7.21: $L = \{ss \mid s \in \{0,1\}^*\}$ (2/6)

z を $uvwxy$ に分割
(ただし $|vwx| \leq n$ かつ $vx \neq \varepsilon$)

観察1: $|uwy| \geq 3n$

- ❖ $|uvwxy| = 4n$ で $|vwx| \leq n$ より $|uy| \geq 3n$
- ❖ $|uwy| \geq |uy| \geq 3n$ を得る



例7.21: $L = \{ss \mid s \in \{0,1\}^*\}$ (3/6)

観察2: $uwy = tt$ の形で書けるはず (Lの語の形より)

- ❖ uwy は 0回の繰返し $uv^0wx^0y = uwy$ に相当
- ❖ 繰返し定理より $uwy \in L$ となるはず

以下では uwy の z 上での位置で場合分け(7通り)

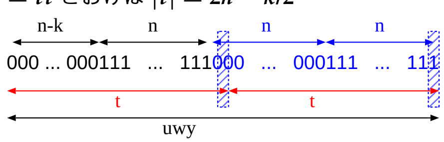
例7.21: $L = \{ss \mid s \in \{0,1\}^*\}$ (4/6)

場合1: vwx が最初の 0^n に含まれる場合

- ❖ $vx = 0^k$ (ただし $0 < k \leq n$) とおく

$uwy = uv^0wx^0y = 0^{n-k}1^n0^n1^n$ を観察

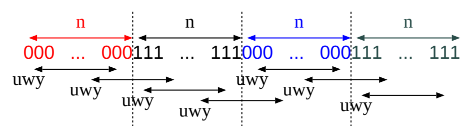
- ❖ $uwy = tt$ とおけば $|t| = 2n - k/2$



- ❖ 最初の t の最後の記号は0
  (最初の t は uwy の最初の1のブロック内で終らない)
- ❖ 2番目の t の最後の記号は1
- ❖ 矛盾 (tt の形をしてない)

例7.21: $L = \{ss \mid s \in \{0,1\}^*\}$ (5/6)

場合(その他): いずれの場合も同様に矛盾が示せる



$z \in L$ のいかなる分割に対しても
繰返しで得られる列は L に含まれない

L は反復補題を満たさない

結論: L は文脈自由言語(CFL)ではない 【証明おわり】

例7.21: $L = \{ss \mid s \in \{0,1\}^*\}$ (6/6)

プログラムコード片1 (正しい)

```
int foo;
foo = 0;
```

プログラムコード片2 (間違い: 未宣言変数の使用)

```
int foo;
bar = 0;
```

$L = \{ss\}$: 変数の宣言(s)と使用(s)の一致を要求

C言語やJavaなどのプログラミング言語がこれに相当

教科書308ページ7.2.1 (b)

おわり