

確定節文法の学習にもとづく輪郭図形の構文的パターン認識

井上 寛樹[†] 中村 克彦[‡]

東京電機大学大学院理工学研究科[†] 東京電機大学理工学部[‡]

1 まえがき

構文的パターン認識は、認識対象を記述した要素の記号列に対して文の構造を形式的に解析する構文解析を用いて図形などのパターンを分析・分類する方式である。

本研究は、確定節文法 (DCG : definite clause grammar) の漸次学習方式 [1] を応用し、輪郭図形を対象に構文的パターン認識を行うことを目的としている。DCG を用いた構文的パターン認識により、図形の要素とその組み合わせに関する数値的な条件を定義することができる。また、漸次学習の機能によって不完全な図形の記述からのパターン認識および例からの図形パターンの学習が可能である。

2 確定節文法とその推論

確定節文法 (DCG) は論理プログラミングを基礎とした拡張文脈自由文法であり、各文法規則を Prolog のホーン節規則に変換して構文解析プログラムや文の導出に使われる。DCG の文法規則は次の形式で表される。

$$P \rightarrow Q_1, \dots, Q_k \quad (k \geq 1)$$

ここで、 P 、 Q は p または $p(T)$ となる形式の項である。左辺に現れる記号 P を非終端記号、 T を DCG 項、左辺に現れない記号列 p は終端記号である。DCG 項 T は、規則の適用の条件を与えたり、部分的な構文解析の結果を伝達することができる。

文法推論とは、ある言語に含まれる記号列 (正例) の集合と含まれない記号列 (負例) の集合が与えられたときに、これらを満足する規則集合 (文法) を合成することである。われわれは、文法推論を行う Synapse システムを開発しており、文脈自由文法および確定節文法の自動合成の改良を続けている [1]。Synapse システムはブリッジ法と呼ばれる規則生成方式と被覆もとづく規則集合探索を組み合わせた文法の漸次学習を行うシステムである。

ブリッジ法は構文解析に失敗したときに得られた不完全な導出木にもとづいて欠けた部分を補うように規則を追加する。開始記号から導出木の不足した部分を補うような規則を下向きに探索し、導出木の欠けた部分を橋渡しするような規則を生成し、規則集合に追加する。ブリッジ法は入力された記号列に対して導出木が完成できるような規則を追加するため、効率的な文法合成が実現できる。

3 構文的パターン認識

一般に輪郭図形の構文的パターン認識では、認識対象となる図形を基本要素 (プリミティブ) の系列で表し、輪郭図形を表す記号列の集合を言語に対応させる。このパターン言語に対して文法規則をつくり、構文解析を行うことでパターン进行分类・分析する。

文脈自由文法による構文的パターン認識では、数量の条件を表すことができない。形が似ている輪郭図形を認識する場合、大量のプリミティブと規則を用いる必要がある。

われわれは曲率一定の部分をプリミティブとみなし、 $p(L, C)$ の形式の項で表す。ここで、 L はプリミティブの線分の長さ、 C は曲率を表す (図 1)。曲率と線分の長さから回転角は $\Delta\theta = LC$ によって与えられる。

輪郭図形をこのようなプリミティブの系列で表すことができる。プリミティブは曲率が変化する部分で区切る。閉じた輪郭図形に対してはプリミティブの回転角の合計が 360 度となる。

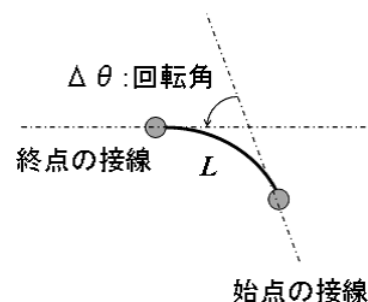


図 1: プリミティブ $p(L, C)$ の回転角

Syntactic Pattern Recognition of Outline Figures Based on Learning of DCG

[†]Hiroki Inoue, [‡]Katsuhiko Nakamura

[†]Graduate School of Science and Engineering, Tokyo Denki University

[‡]School of Science and Engineering, Tokyo Denki University

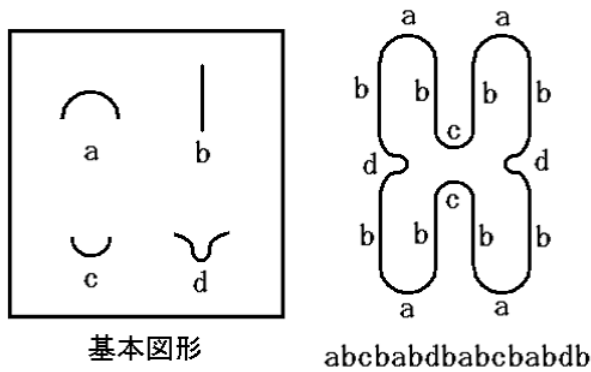


図 2: H 型染色体の輪郭図形 [3]

直線と点は次のプリミティブで表される.

- 長さ L の直線のプリミティブは $line(L, 0)$.
- 曲率が不連続的に変化する点は $point(0, C)$. たとえば直角をなす点は $point(0, \frac{2}{\pi})$ となる.

4 染色体文法の例

構文的パターン認識の代表的な例として, 染色体の形状認識が知られている [2]. 図 2 の輪郭図形は H 型染色体を表している. 基本図形 a, c は曲線のプリミティブ, b は直線のプリミティブ, d は 3 つの曲線のプリミティブで表されている.

H 型染色体のための確定節文法を図 3 に示す. 述語 $chromosome(X0, X4, L, C)$ はプリミティブの差分リスト $X0 - X4$ で表された図形が長さ L , 回転角 $A (= 2\pi)$ の H 型染色体であることを意味している.

この Prolog プログラムに対し, 次の形式の質問を与えることによって図形の認識処理が行われる.

$? - chromosome([P_1, \dots, P_n], [], L, 2\pi).$

ここで, $[P_1 \dots P_n]$ は実際のプリミティブの系列を表すリストである.

この方式にもとづいた DCG の推論によって, 欠けた部分を補うような認識や図形パタンの学習が可能になる.

5 むすび

DCG にもとづく輪郭図形の構文的パターン認識について述べた. 本方式により, 角度や長さなどの量的な条件を与えることができる.

今後の課題として, 輪郭図形のパターンを増やし, パターン認識の条件を検討・評価すること, 一部の欠けた輪郭図形のパターン認識を行うこと, 輪郭図形のパタンの学習が挙げられる.

%H 型染色体

```
chromosome(X0,X4,L,A):-
```

```
  chp(X0,X1,L1,A1), dip(X1,X2,L2,A2),
  chp(X2,X3,L3,A3), dip(X3,X4,L4,A4),
  L is L1+L2+L3+L4, A is A1+A2+A3+A4,
  A > 1.9*pi, A < 2.1*pi.
```

%横のくぼみを除いた上半分と下半分の部分

```
chp(X0,X3,L,A):-
```

```
  rod(X0,X1,L1,A1), bottom(X1,X2,L2,A2),
  rod(X2,X3,L3,A3), L is L1+L2+L3,
  A is A1+A2+A3, A > 0.8*pi, A < 1.2*pi.
```

%左上・右上・左下・右下の U 字の部分

```
rod(X0,X3,L,A):-
```

```
  line(X0,X1,L1,A1), arc(X1,X2,L2,A2),
  line(X2,X3,L3,A3),
  L is L1+L2+L3, A is A1+A2+A3,
  A > 0.8*pi, A < 1.2*pi.
```

%縦のくぼみの部分

```
bottom(X0,X1,L,A):-
```

```
  arc(X0,X1,L,A), A < -0.8*pi, A > -1.2*pi.
```

%横のくぼみの部分

```
dip(X0,X3,L,A):-
```

```
  arc(X0,X1,L1,A1), A1 > 0.2*pi,
  arc(X1,X2,L2,A2), A2 < -0.5*pi,
  arc(X2,X3,L3,A3), A3 > 0.2*pi,
  L is L1 +L2+L3.
  A is A1+A2+A3, A > -0.1*pi, A < 0.1*pi,
```

%直線と曲線 (プリミティブ)

```
line([line(L,A)|X],X,L,A).
```

```
arc([arc(L,A)|X],X,L,A).
```

図 3: H 型染色体のための確定節文法

参考文献

- [1] Keita Imada, Katsuhiko Nakamura: Search for Minimal and Semi-Minimal Rule Sets in Incremental Learning of Context-Free and Definite Clause Grammars. IEICE Transactions 93-D(5): 1197-1204 (2010)
- [2] King.Sun.Fu, Syntactic Pattern recognition Applications, PRENTICE-HALL, INC, 1982
- [3] 山田朋弘, 確定節文法および文法推論の構文的パターン認識への応用, 修士論文, 2009