

# グラフ言語理論のパターン認識への応用

## — 甲状腺画像の認識 —

Application of graph language theory to pattern recognition

日本大学文理学部応用数学科

夜久研究室 齋藤有紀

Yuki SAITOH

**あらまし** コンピュータの普及に伴って認識システムの開発がますます大きな課題になっている。この論文では、甲状腺画像をモデルとした言語理論的アプローチによるパターン認識の研究を行う。最終目標は甲状腺の形状を認識し、正常か異常かを判別する事である。このシステムは入力画像を符号化した後、それを木表現に変換し、構文解析を行う事によってそれを実現するという方法をとっている。

**キーワード** グラフ, 言語理論, パターン認識, 甲状腺

## 1 はじめに

現在、パターン認識を対象とした研究は各方面で実用化している。パターン認識とは種々の対象をあらかじめ想定されたパターンのいずれかを判定する処理をいう。対象として考えられるものに図形や文字・音声などがあげられる。パターン認識の研究のアプローチとして、大別すると2つの方法がある。即ち解析学的アプローチと言語理論アプローチがある。前者は、未知のパターンから特徴抽出を行い、識別関数を当てはめることによって、種々のカテゴリーに分類する方法である。後者は言語理論でいう文法をカテゴリーごとに用意し、未知のパターンの特徴系列がこの文法に適合するか否かで、その属性を決定する方法である。

今回は言語理論的アプローチを取り扱う。特にグラフ言語を理論を用いて、甲状腺画像をモデルとする。最終目標は甲状腺の形状を認識し、正常か異常かを判別する事である。このシステムは入力画像を符号化した後、それを木表現に変換し、構文解析を行う事によってそれ

を実現するという方法をとっている。この研究は、甲状腺の異常を発見することに役立つ。誰もが持っている体の一部であり、早く正確に異常を発見して、一人でも多くの人の役に立つ事を目的とする。

グラフ … ネットワークの数学的モデル  
言語理論 … 文法により文字や図形の集合を解析する理論

パターン認識 … 種々図形・文字・音声をあらかじめ想定されたパターンのいずれかを判定する

甲状腺 … 内分泌腺の1つ。のどぼとけの下方に位置し、気管の両側にある。甲状腺ホルモンを分泌して代謝を調節する。外見は赤みの強い桃色で蝶のような形をしている。

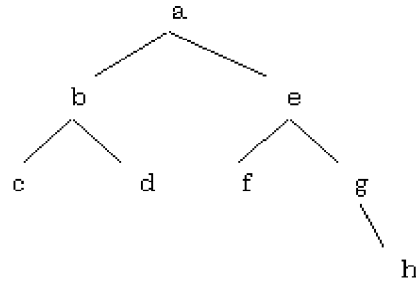
## 2 準備

まず初めに、甲状腺画像認識処理を行う際に必要な木文法について

**定義 2.1 木領域の定義**

木領域  $D$  は、以下に示す (1) ~ (4) の条件を満たす集合である。

- (1)  $1 \in D$  であり、 $1$  は木の根である。
- (2)  $a \in D$  のとき、 $a$  に直接つづく節点 (子) は  $a.i$  の形式である。ここで  $i$  は自然数  $N$  の要素である。
- (3)  $a.i \in D$  のとき、 $a \in D$  かつ  $a.1, a.2 \cdots a.i \in D$  である。
- (4)  $a.i, a.j \in D$  のとき、節点  $a.j$  が節点  $a.i$  の右にあれば、 $i < j$  である。



**定義 2.3 木文法の定義**

木文法  $G_t$  は、4 字組  $(V_N, V_T, P, S)$  である。ここで、

- $V_N$  : 非終端記号の有限集合
- $V_T$  : 終端記号の有限集合
- $P$  : 書き換え規則の有限集合
- $S$  : 開始木の有限集合

である。

木  $T_1$  が  $G_t$  の規則  $P$  により書き換えられて  $T_2$  となるとき、 $T_1 \Rightarrow T_2$  と書く。また  $T_I \Rightarrow T_1, T_1 \Rightarrow T_2, T_2 \Rightarrow T_3, \dots, T_n \Rightarrow T_{II}$  のとき  $T_I \Rightarrow T_{II}$  と書く。

**定義 2.4 木言語の定義**

木文法  $G_t$  により生成される木言語とは、 $L(G_t) = \{T \mid T = (D, L), \forall a \in D, L(a) \in V_T \text{ かつ } \exists T_i \in S, T_i \Rightarrow T\}$  なる集合である。

**定義 2.2 木の定義**

木  $T$  は、 $(D, L)$  の 2 字組である。ここで  $D$  は木領域であり、 $L$  は  $D \rightarrow \Sigma$  のラベルの関数である。ここで  $\Sigma$  はラベルの集合である。

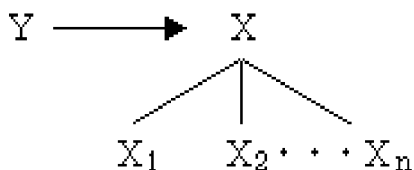
木  $T$  の例

木  $T = (D, L)$  を次のそれぞれの式により定めると次のようになる。

- $D = \{1, 1.1, 1.2, 1.1.1, 1.1.2, 1.2.1, 1.2.2, 1.2.2.1\}$
- $L = \{(1,a), (1.1,b), (1.2,e), (1.1.1,c), (1.1.2,d), (1.2.1,f), (1.2.2,g), (1.2.2.1,h)\}$

**定義 2.5 正規木文法の定義**

木文法  $G_t$  は、以下の条件を満たすとき正規木文法である。開始木の集合  $S$  は、節点数 1 の木の有限集合である。アルファベット  $V_N \cup V_T$  はランク 付きのアルファベットであり  $(V_N \cup V_T, r)$  と書ける。ここで  $r$  は  $V_N \cup V_T$  から非負関数の有限集合への関数である。ある記号のランクは、その記号をラベルとして持つ節点の子の数を示す。規則  $P$  は図 1 に示す形式であり、 $Y, X_1, X_2, \dots, X_n \in V_N, X \in V_T, n \in r(X)$  である。

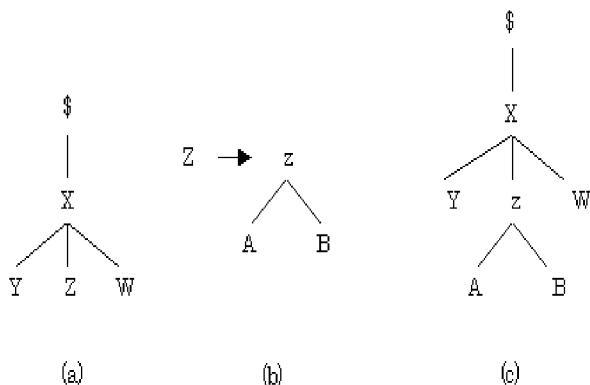


正規木文法の規則P

図1の規則を用いて木  $T = (D, L)$  の葉  $a$  が書き換えられたとする.  $T \Rightarrow T'$  で  $T' = (D', L')$  であり, 節点  $a$  が書き換えられるので,  $L(a) = Y$  である. このとき  $D' = D \cup \{a.i \mid i = 1, 2, \dots, n\}$  である. また,  $L'$  は  $L'(a) = X$ ,  $L'(a.i) = X_i$  となる.

正規木文法による書き換えの例

(a) の木  $T$  が与えられ, (b) の規則により書き換えを行うと (c) になる.



### 定義 2.6 正規木文法の定義

自由木文法は, その規則が以下に示す3種類の形式のいずれかである木文法である.

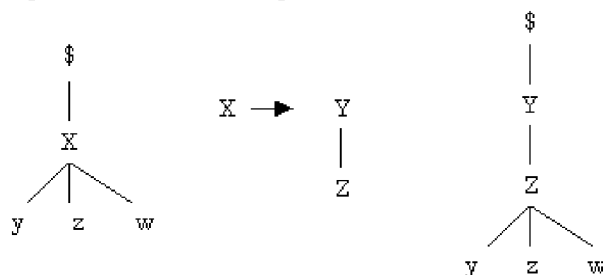
- (1)  $X \rightarrow YZ, 1$
- (2)  $X \rightarrow YZ, 2$
- (3)  $X \rightarrow x, 3$

ここで,  $X, Y, Z \in V_N$  である.

今後 (1),(2),(3) の形式の書き換え規則をそれぞれタイプ1, タイプ2, タイプ3と呼ぶ

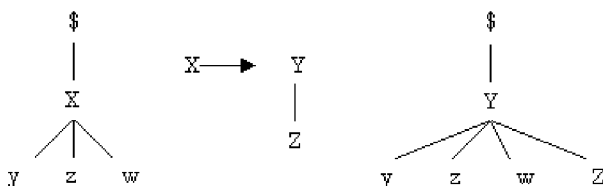
タイプ1 ( $X \rightarrow YZ, 1$ )

木  $T_1 = (D_1, L_1)$  の  $L_1(a) = X$  なる節点  $a$  が書き換えられ,  $T_1 \Rightarrow T'_1$  であるとする. このとき,  $T'_1 = (D'_1, L'_1)$  の  $D'$  は  $\{a.1\} \cup \{a.1.c \mid a.c \in D\} \cup \{D_1 - \{a.c \mid a.c \in D_1\}\}$  となる.



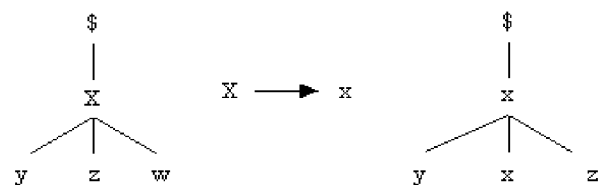
タイプ2 ( $X \rightarrow YZ, 2$ )

木  $T_2 = (D_2, L_2)$  の  $L_2(a) = X$  なる節点  $a$  が書き換えられ  $T_2 = T'_2$  であるとする. このとき,  $k = \max \{j \mid a.j \in D_2\}$  であるとすると  $D'_2 = D_2 \cup \{a.(k+1)\}$  である.



タイプ3 ( $X \rightarrow x, 3$ )

木  $T_3 = (D_3, L_3)$  の  $L_3(a) = X$  なる節点  $a$  が書き換えられ  $T_3 = T'_3$  であるとする. このとき,  $D_3 = D'_3$  である.



### 3 甲状腺画像認識への応用

ここでは甲状腺の形状を識別することを目標とする。甲状腺の画像診断法として頸部軟X線撮影，超音波検査，甲状腺シンチスキャン，CTスキャンやリンパ管造影などがある。

甲状腺とは，のどぼとけの下方に位置し，気管の両側にある。内分泌腺の1つであり，甲状腺ホルモンを分泌して代謝を調節する。外見は赤みの強い桃色で蝶のような形をしていて，左右の葉で気管の上部を前方より覆って存在している。目方はおよそ25グラムである。

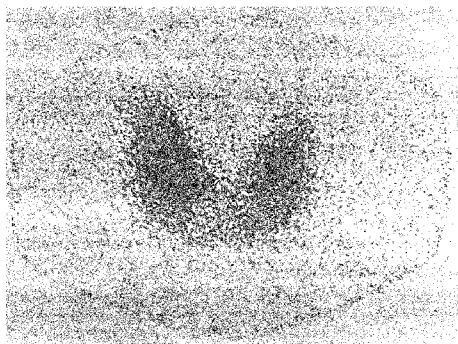


図1. 甲状腺シンチスキャンによる甲状腺画像

甲状腺に異常があると形状に変化が現われる。甲状腺の輪郭部分に凹凸が現われたり，肥大したり，分裂したりする。甲状腺画像処理として，20×20のサイズの雑音のない小型・理想化された2値画像を用いる。そして，次の手順に従って甲状腺の形状を識別する。

#### 3.1 画像の形状コード化

図1で与えられた画像から2値画像に変換された甲状腺画像の各々の交点を表1に従って変換していく。すると図2のようになる。ここで，表1にある6，9の輪郭に対してコードが“\*”となっている。これは正常な甲状腺ではあり得ないコードなので，6，9の輪郭に当てはまるもの(図3)は異常であるとみなす。このコードに当てはまると「接触」という異常メッセージを出力して画像認識処理を終了する。

	境界の形状	形状コード		境界の形状	形状コード
1		a	8		g
2		b	9		*
3		c	10		h
4		d	11		i
5		e	12		j
6		*	13		k
7		f	14		l

表1. 輪郭の形状と対応するコード

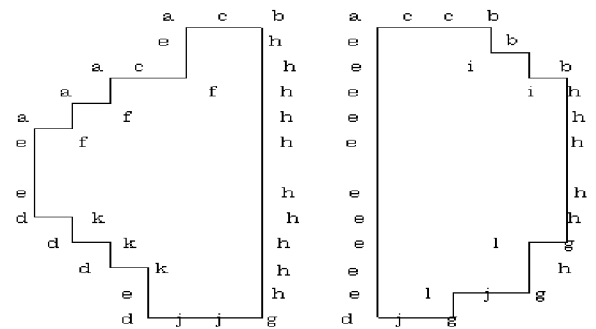


図2. 形状コード化された画像

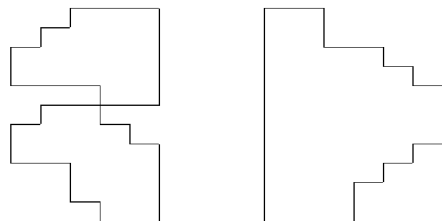


図3. 接触した甲状腺画像

#### 3.2 画像のチェーンコード表現

次の手順として，画像のチェーンコード表現への変換を行う。形状コード化されたコード系列(図2)をもとに，図4のチェーンコードに従って変換する。このとき，左右の葉の始点はそれぞれ画像の左上端で，終点はそれぞれ右下端とする。すなわち，図2の画像が図5に変換される。

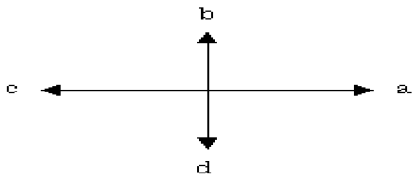


図4. チェーンコード

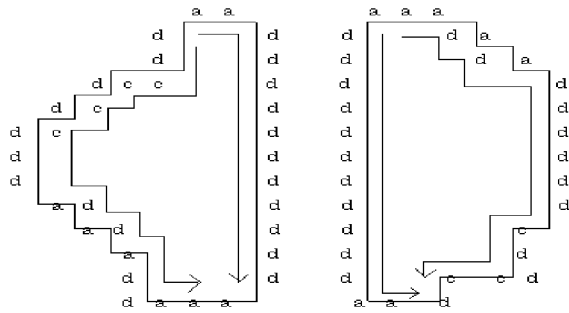


図5. チェーンコード表現

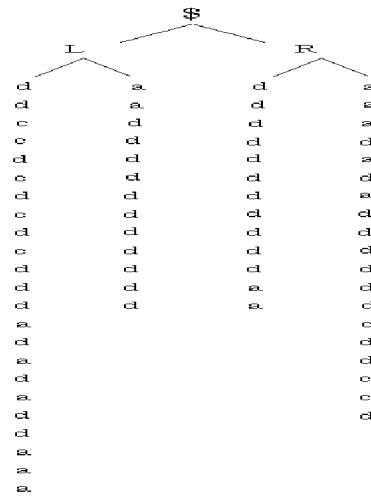


図6. チェーンコードの木表現

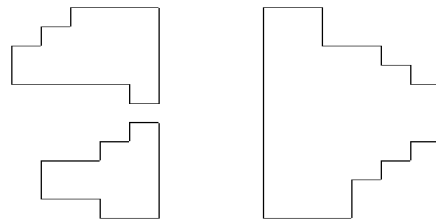


図7. 分裂した甲状腺画像

### 3.3 木表現への変換

そして、作成したチェーンコード表現（図5）を木表現に変換する。木の根は常にラベル\$をもち、正常な甲状腺の場合、左右の像に対応して2つの子節点をもつ。そして、それぞれをラベル”L”および”R”とする。それらに対応する子節点の各々はチェーンコードに従ってさらに子節点を生成し、各々にラベルとしてコードを対応させていく。

すなわち、図5の画像が図6に変換される。また、ここで根の子節点が2つでない場合、正常な甲状腺とみなされない。その甲状腺は左右の葉が分裂している（図7）ことを示している。その時、システムは「分裂」というメッセージを出力して画像認識処理を終了する。

### 3.4 構文解析の手続き

最後に、作成した木に対してあらかじめ用意された正常な形状の甲状腺の木表現を生成する自由木文法  $G_t = (V_N, V_T, P, S)$  をもとに構文解析手続きを適用する。

ここで、 $V_N$  は非終端記号の集合、 $V_T$  は終端記号の集合、 $P$  は書き換え規則の集合、 $S$  は開始木の集合である。ここでの処理は入力する木が構文解析手続きによって  $G_t$  の文であるかどうかを判断し、それによって最初に入力された甲状腺の画像が正常な形状であるかを判定する。

$$G_t = (V_N, V_T, P, S)$$

$$V_N = \{S, S'', [L], L'', [R], R'', [LL], [LL]', [LL]'', [LR], [RL], [RR], [RR]', [RR]'', [A], [B], [C], [D], [E], [E]'\}$$

,[E]'' , [F] , A' , C' , D' , \$' , L' , R' , A , C , D }  
 $V_T = \{ \$ , L , R , a , c , d \}$   
 $P = \{ S \rightarrow S'[L] , 1 \quad [L] \rightarrow L''[LL] , 1 \quad [LL] \rightarrow [LL]'[F] , 1$   
 $[LL]'' \rightarrow [LL]'[B] , 1 \quad [LL]' \rightarrow [A][E]'' , 1 \quad [LL]' \rightarrow [A][LL]'' , 1$   
 $[E]'' \rightarrow [E][B] , 1 \quad L'' \rightarrow L'[LR] , 2 \quad [LR] \rightarrow [F][E] , 1$   
 $S'' \rightarrow S'[R] , 2 \quad [R] \rightarrow R''[RL] , 1 \quad [RL] \rightarrow [E][F] , 1$   
 $R'' \rightarrow R'[RR] , 2 \quad [RR] \rightarrow [F][RR]' , 1 \quad [RR]' \rightarrow [C][RR]'' , 1$   
 $[RR]' \rightarrow [C][E]' , 1 \quad [RR]'' \rightarrow [RR]''[D] , 1$   
 $[E]' \rightarrow [E][D] , 1 \quad [A] \rightarrow D'C' , 1 \quad [B] \rightarrow A'D' , 1$   
 $[C] \rightarrow D'A' , 1 \quad [D] \rightarrow C'D' , 1 \quad [E] \rightarrow [E][E] , 1$   
 $[E] \rightarrow d , 3 \quad [F] \rightarrow [F][E] , 1 \quad [F] \rightarrow a , 3$   
 $A' \rightarrow AA , 1 \quad A' \rightarrow a , 3 \quad C' \rightarrow c , 3 \quad C' \rightarrow CC , 1$   
 $D' \rightarrow DD , 1 \quad D' \rightarrow d , 3 \quad S' \rightarrow \$ , 3 \quad L' \rightarrow L , 3$   
 $R' \rightarrow R , 3 \quad A \rightarrow a , 3 \quad C \rightarrow c , 3 \quad D \rightarrow d , 3 \}$

ここで、入力された木が文法  $G_t$  によって生成されないことが判明すると、システムは「形状異常」とみなされる。また、 $G_t$  によって生成されれば「正常」とみなされる。

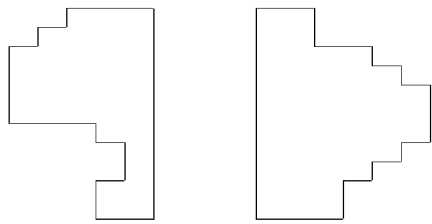
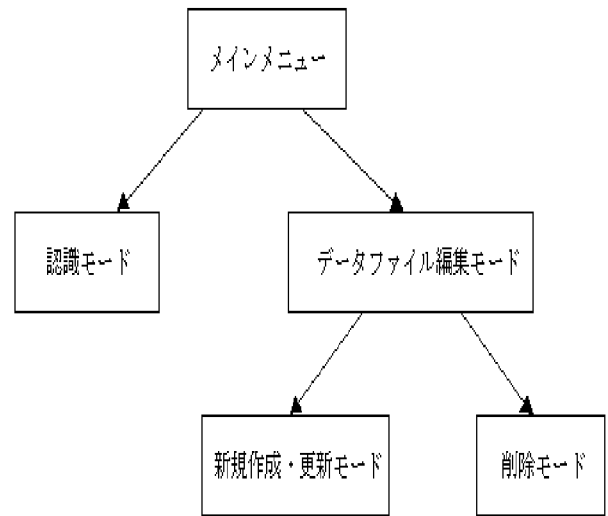


図 9. 凹凸のある異常な甲状腺画像

## 4 JAVA 言語による甲状腺画像認識システムの作成

プログラミングの手順



まず、認識モードとデータファイル編集モードの選択を行う

### 認識モード

①～④の順番で作業を行う。

- ① 形状コード化  
形状コード表にあてはめる
  - ② 画像のチェーンコード表現  
チェーンコードに従って変換
  - ③ 木表現への変換  
2つの子節点をもつ木表現に変換
  - ④ 構文解析の手続き
    - ①で\*が存在すれば、接触
    - ③で変換できなければ、分裂
    - ④で受理されなければ、凹凸または肥大
- ①～④全てに当てはまれば、正常

### データファイル編集モード

新規作成・更新モードと削除モードの選択

### 新規作成・更新モード

ファイルの新規作成、更新

データの編集  
データの Save

### 削除モード

ファイルの削除を行う

## 5 結論

このシステムでは、正常な形状をもつ甲状腺及び異常な形状をもつ甲状腺いずれにおいても識別できるという事が分かった。

## 6 おわりに

現在の段階では、正常な甲状腺と接触や分裂などの異常な甲状腺とを識別するのみの処理にとどまっている。本来ならば異常な甲状腺に対して適切な処理を行うことが必要となる。今後の課題としては、その後の引き続きの適切な処理、アルゴリズムの効率化による処理速度の向上、画素数の増加による特徴抽出の問題、他の手法との比較、システムの信頼性を高めるためのより多くの甲状腺シンチグラムによるテストを行うことである。

## 参考文献

- [1] 添田明宏, 甲状腺画像認識システム, FARST ソフトウェア報告書 (1985)
- [2] 細田英和, 「グラフ言語理論のパターン認識への応用」, 日本大学文理学部応用数学科卒業研究報告書 (1998)
- [3] 南昌樹, 「グラフ言語理論のパターン認識への応用」, 日本大学文理学部応用数学科卒業研究報告書 (1999)
- [4] 河西朝雄, JAVA 入門, 技術評論社, 1996, 東京都, 415

- [5] 基礎から学ぶ Java プログラミング, 電子開発学園出版局, 2000, 東京都, 373
- [6] 丸の内とら, 初体験 Java, 技術評論社, 2000, 東京都, 295
- [7] 結城浩, Java 言語プログラミングレッスン上, SOFTBANK, 1999, 東京都, 354
- [8] 結城浩, Java 言語プログラミングレッスン下, SOFTBANK, 1999, 東京都, 319
- [9] はじめて学ぶ PASCAL, 共立出版株式会社, 1990, 東京都, 169
- [10] 真野芳久, Pascal プログラミングの基礎, サイエンス社, 1994, 東京都, 265
- [11] 和賀井敏夫, 佐久間浩, 乳房・甲状腺アトラス, ベクトルコア, 1999, 東京都, 169