Vol. 36 No. 2　　情　報　処　理　　Feb. 1995

解　説

確　率　文　法†

日　高　　達††

1. はじめに

最近，事例(解析)データを活用することにより自然言語処理の質の向上を計る方法が模索されている．事例データの文に類似した入力文に遭遇したとき，その事例データの正しい解析(事例解析データ)に類似する解析を優先しようとする考えに基づいている．事例データを言語処理に反映する方法は，古くは確率有限オートマトン(Finite State Automaton)としてモデル化され文字認識や音声認識に応用されてきたし，最近では確率有限オートマトンの自然な拡張として，確率文脈自由文法(Probabilistic Context Free Grammar)がテキスト認識や文の音声認識における数学モデルとして広く用いられている．テキスト認識や文の音声認識では，文字図形，部分音声入力から，それぞれ文字，単語を推定する要素認識の段階と，要素列から文または構文を推定する言語認識の段階がある．要素認識の段階では，多くの場合，パターン入力の特徴抽出結果に基づき，文字または単語を確率的に推定する手法がとられる．この場合，確率モデルが採用されるが，確率モデルはいくつかのパラメタを含んでおり，標本(事例データ)に依存してパラメタが設定される．概略的には，標本集合が最も発生しやすいように，パラメタを設定する方法がとられる．したがって，標本の統計的な特徴が確率モデルに反映されることになる．次に処理は言語認識に移っていくのであるが，前段が確率モデルを採用しているので，後段の言語認識段階でも確率モデルを採用するほうが整合性が良い．単純な例として手書き文字認識の場合を考えてみよう．

文字図形はあらかじめ設定された桝目の中に一つずつ書かれるものとし，n個の桝目に書かれた文字図形列を$x_1 \cdots x_n$で表す．入力文字図形列が$x_1 \cdots x_n$のときに，これを書いた人が意図した文字列が$c_1 \cdots c_n$である条件付き確率を$P_r(c_1 \cdots c_n | x_1 \cdots x_n)$で表す．文字認識が，$P_r(c_1 \cdots c_n | x_1 \cdots x_n)$を最大にする$c_1 \cdots c_n$を推定することを目的にする場合，ベイズの定理により，

$$P_r(c_1 \cdots c_n | x_1 \cdots x_n) = \frac{P_r(c_1 \cdots c_n) \cdot P_r(x_1 \cdots x_n | c_1 \cdots c_n)}{P_r(x_1 \cdots x_n)} \quad (1)$$

ここで，$P_r(c_1 \cdots c_n)$，$P_r(x_1 \cdots x_n)$，$P_r(x_1 \cdots x_n | c_1 \cdots c_n)$は，それぞれ文$c_1 \cdots c_n$が発生する確率，文字図形列$x_1 \cdots x_n$が発生する確率，文字列$c_1 \cdots c_n$が意図されたときに文字図形列$x_1 \cdots x_n$として書かれる条件付き確率である．各文字図形は桝目の中に一つずつ丁寧に書かれ，したがって直前の文字図形の形状が直後の文字図形の形状に影響しない場合，近似的に次式が成立する．

$$P_r(x_1 \cdots x_n | c_1 \cdots c_n) = \prod_{i=1}^{n} P_r(x_i | c_i) \quad (2)$$

したがって，(1)式を最大にする文字列の推定は，(3)式を最大にする文字列$c_1 \cdots c_n$を推定することと同値である．

$$P_r(c_1 \cdots c_n) \cdot \prod_{i=1}^{n} P_r(x_i | c_i) \quad (3)$$

文字認識では(1)式の値を最大にする$c_1 \cdots c_n$を求めることで満足することも考えられるが，文字(音声)認識がヒューマンインタフェースの一環として行われる場合にはこれでは不満足であり，文の構文構造Tを求める．すなわち，$P_r(T | x_1 \cdots x_n)$を最大にする構文構造Tを推定する必要がある．(1)式を最大にする文$c_1 \cdots c_n$が求められても，その構文構造には曖昧さが存在し，したがって$P_r(T | x_1 \cdots x_n)$を最大にす

† Probabilistic Grammar by Toru HITAKA(Department of Computer Science and Communication Engineering, Faculty of Engineering, Kyushu University).

†† 九州大学工学部情報工学科

る T が $c_1 \cdot \cdots \cdot c_n$ の構文構造とは限らないからである．

$$P_r(T|x_1 \cdot \cdots \cdot x_n)=\frac{P_r(T)\cdot P_r(x_1 \cdot \cdots \cdot x_n|T)}{P_r(x_1 \cdot \cdots \cdot x_n)} \quad (4)$$

だから，(4)式を最大にする T は，(5)を最大にする T である．

$$P_r(T)\cdot P_r(x_1 \cdot \cdots \cdot x_n|T) \quad (5)$$

(5)において，$P_r(T)$ は構文構造 T が生起する確率である．T を構文構造とする文(T の葉の列)を $c_1 \cdot \cdots \cdot c_n(T)$ とすると，次式が成立する．

$$P_r(x_1 \cdot \cdots \cdot x_n|T)=P_r(x_1 \cdot \cdots \cdot x_n|c_1 \cdot \cdots \cdot c_n(T))=\prod_{i=1}^{n} P_r(x_i|c_i(T)) \quad (6)$$

(3)，(6)式において，$P(x_i|c_i)$ はいわゆる一文字認識の技術で推定される値であるが，$P_r(c_1 \cdot \cdots \cdot c_n)$，$P_r(T)$ は，それぞれ文(としての文字列) $c_1 \cdot \cdots \cdot c_n$，構文構造 T が発生する確率である．したがって，$P_r(c_1 \cdot \cdots \cdot c_n)$ や $P_r(T)$ を計算するためには，文の生成機構(文法)を確率化する必要があるのである．また，自然言語処理においては入力文 $c_1 \cdot \cdots \cdot c_n$ に対しその構文構造 T を推定する必要があるが，一般的に自然言語文の文法構造は曖昧である(一つの文に対し複数の構文構造が対応する)から，曖昧さを解消する一つの手立てとして，事例解析データ(標本として収集された構文構造 $T_1, T_2, \cdots, T_N$) の生起確率を高くするような確率文法を導入し，$P_r(T|c_1 \cdot \cdots \cdot c_n)$ を最大にする構文構造 T を求めることにより，曖昧性の解消を計ることが考えられる．

本稿では，確率文脈自由文法の定義と統計的性質，認定問題と構文解析法およびその能率，確率パラメタの推定法，確率文法のパターン認識における応用の概略を紹介し，特に，事例データの言語処理への利用法に関して重要な確率パラメタの推定法を詳しく述べる．

本稿では，読者は文脈自由文法に関する基礎的知識はもっていることを前提とする．

2. 確率文脈自由文法理論

確率文脈自由文法の定義，構文解析法，パラメタ設定法について解説する．

2.1 確率文脈自由文法の定義

【定義 1】 確率文脈自由文法 PCFG G は，次のような4組で定義される．

$G=(\Sigma, V, P, S, p)$

Σ：終端記号(a, b などの小文字で表す)の有限集合

V：非終端記号(X, Y などの大文字で表す)の有限集合

S：開始記号($S \in V$)

P：$\{X \to \alpha | X \in V, \alpha \in (\Sigma \vee V)^*\}$ の有限部分集合

p：$P \to (0, 1]$　(P から区間 $(0, 1]$ への写像)

P の要素は CFG における書換え規則である．$X \to \alpha \in P$ と値 $p(X \to \alpha)$ の組を

$$X \xrightarrow[p(X\to\alpha)]{} \alpha$$

で表し，(確率付き)書換え規則と言う．また，$X, \alpha, p(X \to \alpha)$ を上の書換え規則のそれぞれ左辺，右辺，適用確率と言う．左辺が同一の非終端記号(X とする)であるすべての書換え規則

$$X \xrightarrow[p_i]{} \alpha_i \quad (i=1, 2, \cdots, I_X) \quad (7)$$

に対し，適用確率の総和は1になるように設定する．

$$p_1+p_2+\cdots+p_{I_X}=1. \quad (8)$$

本稿で扱う PCFG は無駄な非終端記号を含まない，すなわち，任意の $X \in V$ に対して，$\omega \in \Sigma^*$ と $\alpha \in (\Sigma \cup V)^*$，$\beta \in (\Sigma \cup V)^*$ が存在し，

$$X \overset{+}{\Rightarrow} \omega,$$

$$S \overset{*}{\Rightarrow} \alpha X \beta$$

が成立するものとする．

文，言語 $L(G)$，導出(木)，最左(右)導出などの定義は CFG の場合とまったく同じなので，詳しい定義は省略する．本稿では，X を root とし，すべての葉が終端記号であるような導出木を，特に，X からの構文木と呼び，葉が非終端記号であってもよい導出木と区別することにする．開始記号 S からの構文木を，簡略に，構文木と呼ぶ．PCFG では文 $\boldsymbol{s} \in \boldsymbol{L(G)}$, $\boldsymbol{s}$ の構文木 T に対し，それぞれ s, T の生起確率 $\boldsymbol{P_r(s)}$, $\boldsymbol{P_r(T)}$ が次のように定義される．すなわち，$\boldsymbol{P_r(s)}$ は文 s を生成するすべての構文木の生起確率の総和であり，T の生起確率 $\boldsymbol{P_r(T)}$ は T の導出に適用された書換え規則の適用確率の(重複を含めた)積である．CFG における文 s の認定($\boldsymbol{s} \in \boldsymbol{L(G)}$の判定)と構文解析($s$ の構文木を求めること)に対して，PCFG では，文 s の生起確率 $\boldsymbol{P_r(s)}$ を求めるこ

とが**認定問題**であり，文 s を生成する構文木の中で生起確率の最も大きい構文木を求めることが**構文解析**である．□

PCFG の書換え規則 $X \underset{p}{\rightarrow} \alpha$ の意味はおよそ次のようなものである．すなわち，文の発生過程を文脈自由文法による確率的生成過程をみるとき，生成の途中の文形式に出現した非終端記号 X は，次に X を左辺にもつ書換え規則の適用を受けるわけであるが，その場合，$X \rightarrow \alpha$ の適用を受けて α に書き換わる確率が p である．

自然言語の PCFG では，p が $p : P \times P \rightarrow (0, 1]$ で定義される場合がある．この場合の p は，$X \rightarrow \alpha \in P$ が $Y \rightarrow \beta \in P$ で生成された β の中の X に適用されるときの(条件付き)適用確率が $p(X \rightarrow \alpha, Y \rightarrow \beta)$ であることを意味する．この形式の文法を **Conditional PCFG** と呼ぶことにする．Conditional PCFG は標準形の PCFG に等価変換できる(**付録**参照)．変換アルゴリズムから分かるように，等価変換された標準形の PCFG では，元の Conditional PCFG に比べて非終端記号の個数と書換え規則の個数が膨大になり，文法の管理や見通しが悪くなる．

Conditional PCFG の変種はほかにも種々考えられるが，いずれも標準形の PCFG に等価変換可能と考えてよい．Conditional PCFG は自然言語の文法にしばしば用いられ，**4.**で紹介する確率木接合文法でも類似の形式が用いられている．

2.2 PCFG の構文解析法

PCFG の構文解析は，任意に与えられた文字列 s に対し，最も生起確率の高い構文木を求めることである．PCFG の構文解析は CFG の横型構文解析法 (CYK 法，Earley 法とその亜流のチャート法)を拡張することによって容易に構成することができる．拡張のやり方はほぼ同様なので，ここでは CYK 法の場合を例にとって説明する．

記述を簡潔にするために，まず記号列における positioning の定義から始めよう．

【定義 2】　n 長さ記号列 $s = a_1 \cdot a_2 \cdot \cdots \cdot a_n$ において，a_1 の左隣の位置を position 0，a_i と a_{i+1} の間の位置を position i，a_n の右隣の位置を position n とする．position i から position $j (0 \leq i \leq j \leq n)$ の間の s の部分列を $s(i, j)$ で表す．すなわち，

$$s(i, j) = a_{i+1} \cdot \cdots \cdot a_j.$$

□

CYK 法の概要を簡単に述べておこう．CYK 法は，$X \rightarrow YZ$ や $X \rightarrow \omega (\in \Sigma^*)$ のように，書換え規則の右辺が長さ 2 の非終端記号列であるかまたは長さ 1 以上の終端記号列であるような CFG を対象にしている．CYK 法は **PL 登録ステージ**と**構文木抽出ステージ**からなる．まず，n 長さの入力記号列 $s = a_1 \cdot \cdots \cdot a_n$ に対し，PL 登録ステージでは次に示す 3 変数の述語 $P_G (\subseteq V \times I \times I)$ を true にするような 3 組 (X, i, j) を Parse List PL にすべて登録する．ただし，I は 0〜n の自然数の集合である．

$P_G(X, i, j) = \text{true} \overset{def}{\Longleftrightarrow} s(i, j)$ を leaf とするような X からの構文木が存在する．

第 2 ステージは，$(S, 0, n) \in$PL である場合に，$(S, 0, n)$ が PL に登録される過程を，PL の検索を通して，逆に辿り，S を root とし $s = s(0, n)$ を leaf とする構文木 π を最左導出として出力する．

PCFG G の構文解析では，次のような4変数述語 $P_G (\subseteq V \times I \times I \times (0, 1])$ を用いる．

$P_G(X, i, j, r) = true \overset{def}{\Longleftrightarrow} s(i, j)$ を leaf とするような X からの導出木の中で，最大の生起確率は r である．

PL 登録ステージでは，P_G を true にするようなすべての 4 組が PL に登録される．

【PL 登録アルゴリズム】

step 1.　$X \rightarrow s(i, j) \in P$ となるようなすべての $X \in V$ と $i, j (0 \leq i \leq j \leq n)$ に対して，$(X, i, j, p(X \rightarrow s(i, j))) \in$PL にする．

step 2.　次の操作を，新たに登録する 4 組がなくなるまで，繰り返し実行する．

$X \rightarrow YZ \in P$, $(Y, i, j, r_1) \in$PL, $(Z, j, k, r_2) \in$ PL，かつ，最初の 3 組が X, i, k である 4 組が PL に登録されていない場合には，$(X, i, k, p(X \rightarrow YZ) \cdot r_1 \cdot r_2) \in$PL とし，登録されている場合 ($(X, i, k, r) \in$PL とする)には，$(X, i, k, r)$ を PL から消去して，$(X, i, k, max\{p(X \rightarrow YZ) \cdot r_1 \cdot r_2, r\}) \in$PL とする．

【構文木抽出アルゴリズム】

最初の 3 組が $S, 0, n$ である 4 組が PL に存在しなければ，$s \in L(G)$ でないので，'error' を出力し停止する．存在する ($(S, 0, n, q) \in$PL とする)

ならば，$\pi \leftarrow \varepsilon$にして，次のRoutine $R(S, 0, n, q, \pi)$を再帰的に実行し，Routine $R(S, 0, n, q, \pi)$が返す最左導出πを構文木として出力する．Routine $R(X, i, k, r, \pi)$は$s(i, k)$をleafとし生起確率がrであるような，Xからの構文木πを返す再帰的手続きである．

Routine $R(X, i, k, r, \pi)$

次のa)またはb)が成立する4組をPLから検索(必ず成功する!)する．

$$\text{a)} \begin{cases} X \rightarrow YZ \in P, \\ (Y, i, j, r_1) \in \mathrm{PL} \\ (Z, j, k, r_2) \in \mathrm{PL}, \\ r = p(X \rightarrow YZ) \cdot r_1 \cdot r_2. \end{cases}$$

$$\text{b)} \begin{cases} X \rightarrow s(i, k) \in P, \\ r = p(X \rightarrow s(i, k)). \end{cases}$$

a)が成立する$(Y, i, j, r_1) \in \mathrm{PL}$，$(Z, j, k, r_2) \in \mathrm{PL}$が検索された場合には，$\pi \leftarrow \pi \cdot (X \rightarrow YZ)$にして，まず$R(Y, i, j, r_1, \pi)$を実行し，次に，これが返す値$\pi$に対して$R(Z, j, k, r_2, \pi)$を実行する．また，b)が成立する場合には，$\pi \leftarrow \pi \cdot (X \rightarrow s(i, k))$を返す． □

上記アルゴリズムにおいて，PLに属する4組(X, i, j, r)の実数値rを扱う(定数ステップ)操作が付加している点以外は，CYK法とまったく同じである．また，$(X, i, j, r) \in PL$，$(X, i, j, r') \in PL$ならば，$r = r'$であることを考慮すると，上記のPCFGの構文解析アルゴリズムの能率はCYK法の能率に等しい．

文sの認定問題は，文sを生成するすべての構文木の生起確率の総和を求める問題である．この場合，上のPL登録アルゴリズムにおいて，$(X, i, k, max\{p(X \rightarrow Y \cdot Z) \cdot r_1 \cdot r_2, r\}) \in PL$を$(X, i, k, r + p(X \rightarrow Y \cdot Z) \cdot r_1 \cdot r_2) \in PL$に変更すれば，PL登録アルゴリズムの実行を終了した時点では，PLには次の4変数述語P'_Gをtrueにするすべての4組が登録されることになる．

$P'_G(X, i, j, r) = true \overset{def}{\Longleftrightarrow} s(i, j)$をleafとするようなすべてのXからの構文木の生起確率の総和はrである．

したがって，$s, 0, n$を最初の3組とする4組をPLから検索すれば，4組目の値が，求める$P_r(s)$である．

上記のアルゴリズムにおいて，PL登録アルゴリズムのworst caseの能率は処理時間$O(n^3)$，記憶領域$O(n^2)$であり，構文抽出アルゴリズムのworst caseの能率は処理時間$O(n^2)$，記憶領域$O(n^2)$であり，これはCFGのparserであるCYK法の能率に等しい．

2.3 確率パラメタの推定

PCFG $G = (\Sigma, V, P, S, p)$において，$p : P \rightarrow (0, 1]$は標本(事例データ)集合に依存して定められるパラメタである．

確率パラメタの推定では，統計学の立場から望ましいとされるいくつかの基準がある．このうち，不偏性と一致性をまず紹介する．次に，一般に考えられるパラメタ推定法として，最大総和推定法と最尤推定法について議論する．

【定義3】 標本集合

ランダムに収集された標本(構文木)を$T_1, T_2, \cdots, T_N$とし，$X \rightarrow \alpha \in P$が構文木$T$の導出に適用された回数を$n(T, X \rightarrow \alpha)$で表す．ここで，$N$個の標本の列$\mathcal{F}_N = (T_1, T_2, \cdots, T_N)$が収集される確率$P_r(\mathcal{F}_N)$は次のようになる．

$$P_r(\mathcal{F}_N) = \prod_{k=1}^{N} P_r(T_k) \qquad (9)$$

□

$\mathcal{F}_N = (T_1, T_2, \cdots, T_N)$に基づいて，パラメタ$p(\delta)(\delta \in P)$の推定がなされるが，この推定値を$\hat{p}(\delta|\mathcal{F}_N)$と記す．

【定義4】 不偏性

任意の$N(\geq 1)$に対し，

$$E\{\hat{p}(\delta|\mathcal{F}_N)\} = p(\delta) \qquad (10)$$

であるとき，そのパラメタ推定は不偏推定(Unbiased Estimate)であると言う．また，

$$\lim_{N \rightarrow \infty} E\{\hat{p}(\delta|\mathcal{F}_N)\} = p(\delta) \qquad (11)$$

であるとき，漸近不偏推定(Asymptoticly Unbiased Estimate)と言う．ただし，$E\{\hat{p}(\delta|\mathcal{F}_N)\}$は長さ$N$の標本列空間における推定値の期待値を表す．

不偏性は，パラメタ推定法が是非満足すべき要請では必ずしもないが，不偏性が成立するパラメタ推定に関する理論展開が容易になる局面が多い． □

次の一致性は，パラメタ推定における要請としてきわめて重要である．

【定義5】 一致性

任意の$\delta \in P$と任意の$\varepsilon(>0)$に対し，

$$\lim_{N\to\infty} P_r\{\mathcal{F}_N \| \hat{p}(\delta|\mathcal{F}_N) - p(\delta)| < \varepsilon\} = 1 \quad (12)$$

であるとき，そのパラメタ推定は一致推定(Consistent Estimate)であると言う．ただし，$P_r\{\mathcal{F}_N \| \hat{p}(\delta|\mathcal{F}_N) - p(\delta)| < \varepsilon\}$ は，推定値 $\hat{p}(\delta|\mathcal{F}_N)$ と真値 $p(\delta)$ の差が $\pm\varepsilon$ 以内であるような標本集合 $\mathcal{F}_N$ が収集される確率である．

一致性は，標本の個数が大きければ，推定値が真値にいくらでも近くなることが確率的に保証されることを意味しており，パラメタ推定上きわめて重要な要請と考えられる． □

パラメタの推定では，正係数の多項式に関するL. E. Baumの不等式理論を用いるので，まずこの紹介から始めよう．

【Baum 理論】 [1),2)]

$M = m_1 + \cdots + m_n$ 個の変数 $x_{ij}(i=1, \cdots, n,\ j=1, \cdots, m_i)$ を簡略に x で表す．$f(x)$ は M 個の変数 x に関する正係数の任意の多項式である．x の変域は，次の(13)，(14)式で定義される領域 D とする．

$$x_{ij} \geq 0 \quad (i=1, \cdots, n,\ j=1, \cdots, m_i) \quad (13)$$

$$x_{i1} + \cdots + x_{im_i} = 1 \quad (i=1, \cdots, n) \quad (14)$$

このとき，任意の値 $x \in D$ に対して

$$x'_{ij} = \frac{x_{ij} \cdot \dfrac{\partial f}{\partial x_{ij}}(x)}{\displaystyle\sum_{k=1}^{m_i} x_{ij} \cdot \frac{\partial f}{\partial x_{ik}}(x)} \begin{pmatrix} i=1, \cdots, n, \\ j=1, \cdots, m_i \end{pmatrix} \quad (15)$$

とすると，$x' \in D$ であり，かつ(16)式が成立する．

$$f(x') \geq f(x) \quad (16)$$

また，(16)式が等式となるのは，$x' = x$ のときのみである． □

2.3.1 最大総和推定

最大総和推定は，Baum 理論を用いて，随意に初期設定された書換え適用確率 $p=p_0$ から出発して，$p=p_1$，$p=p_2, \cdots$ と次々に p を変化されていき，$\sum_{k=1}^{N} P_r(T_k)$ の値を極大値に導いていく手法である．

$p(\delta)$ $(\delta \in P)$

を Baum 理論における変数と考えると，

$$f(p) \overset{\text{def}}{=} \sum_{k=1}^{N} P_r(T_k) = \sum_{k=1}^{N} \prod_{\delta \in P} p(\delta)^{n(T_k, \delta)} \quad (17)$$

は，Baum 理論における非負係数の多項式 f になり，また，p は変域 D の上を動く．

【最大総和推定アルゴリズム】

Step 1. 初期値設定

書換え規則の適用確率 p を適当に定め，これを $p=p_0$ とする．

Step 2. 繰返し操作 $(l=0, 1, 2, \cdots)$

(7)の書換え規則に対し，

$$p_{l+1}(X \to \alpha_i) = \frac{\displaystyle\sum_{k=1}^{N} P_r(T_k|p=p_l) \cdot n(T_k, X \to \alpha_i)}{\displaystyle\sum_{j=1}^{I_X} \sum_{k=1}^{N} P_r(T_k|p=p_l) \cdot n(T_k, X \to \alpha_j)} \quad (18)$$

$(i=1, \cdots, I_X)$ とすると，次式が成立する．

$$\sum_{k=1}^{N} P_r(T_k|p=p_l) \leq \sum_{k=1}^{N} P_r(T_k|p=p_{l+1}) \quad (19)$$

ただし，$P_r(T_k|p=p_l)$ は，$p=p_l$ として計算した T_k の生起確率である． □

最大総和推定アルゴリズムにおける繰返し操作は，$\sum_{k=1}^{N} P_r(T_k|p=p_l)$ の値が十分落ち着いてきたところで打ち切り，$\hat{p} \simeq p_l$ とすればよい．

最大総和推定を一口で言えば，標本集合の生起確率を最大(実際には極大)にする方法であり，一見合理的にみえるが，$P_r(T_1) + \cdots + P_r(T_N)$ は大きくなっても，各 $P_r(T_1), \cdots, P_r(T_N)$ の値をおしなべて大きくする方法ではなく，特定の T_k に対して $P_r(T_k) \simeq 0$ となることが起こり得る．しかし，標本として収集された事例データであるからには，その生起確率が極端に小さくなること，すなわち，生起確率が 0 に近い標本点が採集されることは考えにくい．また，統計学上の立場から言えば，文献10)で示されるように，不偏性も一致性も満足しない不十分な推定法なのである．

2.3.2 最尤推定法(Maximum Likelihood Estimate)

標本 $T_1, \cdots, T_N$ が収集される(または発生する)確率(尤度)を最大にする方法を一般に最尤推定と言う．標本の収集が互いに独立に行われると仮定すると，$T_1, \cdots, T_N$ が収集される確率は(11)式になる．

【定理1】 最尤推定式

(11)式を最大にする確率パラメタ p は次式で与えられる．

$$p(X \to \alpha_i) = \frac{\displaystyle\sum_{k=1}^{N} n(T_k, X \to \alpha_i)}{\displaystyle\sum_{k=1}^{N} \sum_{i=1}^{I_X} n(T_k, X \to \alpha_i)} \quad (20)$$

証明：自然対数 $\log x$ は単調増加関数だから，

(11)式を最大にすることと，次式を最大にすることは同等である．

$$h(p)=\sum_{k=1}^{N} logP_r(T_k)=\sum_{k=1}^{N}\sum_{\delta\in P} n(T_k,\delta)\,logp(\delta) \tag{21}$$

また，$X\in V$ を左辺とする書換え規則(7)について，次の拘束条件が付随する．

$$p(X\to\alpha_1)+p(X\to\alpha_2)+\cdots+p(X\to\alpha_{I_X})=1$$

この拘束条件の下に(23)式を最大にする p は，ラグランジェの未定常数法により，次の $g(p,\lambda)$ を極大にする．

$$g=h(p)+\lambda\left(1-\sum_{i=1}^{I_X}p(X\to\alpha_i)\right) \tag{22}$$

$$\left.\begin{aligned}&\frac{\partial g}{\partial p(X\to\alpha_i)}=\sum_{k=1}^{N}\frac{n(T_k,X\to\alpha_i)}{p(X\to\alpha_i)}-\lambda=0\\&\frac{\partial g}{\partial\lambda}=1-\sum_{i=1}^{I_X}p(X\to\alpha_i)=0\end{aligned}\right\} \tag{23}$$

$$\therefore\lambda=\lambda\left(\sum_{i=1}^{I_X}p(X\to\alpha_i)\right)=\sum_{k=1}^{N}\sum_{i=1}^{I_X}n(T_k,X\to\alpha_i),$$

$$p(X\to\alpha_i)=\frac{1}{\lambda}\sum_{k=1}^{N}n(T_k,X\to\alpha_i).$$

これより，(22)式が導かれる．□

最尤推定では，$P_r(T_1)\cdot\cdots\cdot P_r(T_N)$ を最大にするので，特定の $P_r(T_k)$ が極端に小さくなることがない．最大総和推定では，“標本集合が収集される確率”に対する配慮が甘いために，一致性が成立しないとも考えられる．

最尤推定では，文法を少し制限すれば，漸近不偏性と一致性が成立すると著者は考えており，現在その証明を進めている．

3. 確率文法のパターン認識への応用

文字/音声認識では，**1.**で述べたように，自然言語の確率文法と一文字/一単語(または一文節)認識技術が駆使される．この場合，自然言語の確率文法 G と一文字/一単語認識を統合した，文字図形列/音声波時系列を生成するパターン生成確率文法 G^* を構成して，G^* の構文解析により文字/音声認識を行うことを考えると，理論的にすっきりする．パターン生成確率文法 G^* の構成法を，まず文字認識を例にとって解説しよう．

自然言語の PCFG $G=(\Sigma, V, P, S, p)$ が与えられており，入力文字図形列 $x=x_1\cdot\cdots\cdot x_n$ に対する一文字認識により，すべての $c\in\Sigma$ と各 x_i に対し，$P_r(x_i|c)$ が計算されているとする．$P_r(x_i|c)$ は書き手がカテゴリ c の文字を書くことを意図したとき，文字図形 x_i を書く条件付確率(密度)である．G における書換え規則の適用確率 $p(X\to\alpha)$ は，統語範疇 X の句を生成するとき，最初に適用される統語規則が $X\to\alpha$ である条件付確率であることを考えると，パターン生成確率文法 $G^*(x)$ は次のように構成すればよい．

[**パターン生成確率文法** $G^*(x)$]

$G=(\Sigma, V, P, S, p)$

$G^*(x)=(\Sigma', V', P', S', p')$

$\Sigma'=\{x_1, x_2, \cdots, x_n\}$

$V'=\Sigma\cup V$

$P'=P\cup\{c\to x_i|c\in\Sigma, x_i\in\Sigma'\}$

p'；$X\to\alpha\in P$ に対して，

$p'(X\to\alpha)=p(X\to\alpha)$.

$c\in\Sigma, x_i\in\Sigma'$ に対して，

$p'(c\to x_i)=P_r(x_i|c)$.

□

上記のパターン生成確率文法 $G^*(x)$ では，$c\in\Sigma$ から終端記号(入力文字図形)の書換え規則に対して，

$$p'(c\to x_1)+p'(c\to x_2)+\cdots+p'(c\to x_n)<1$$

となり，確率文法の定義における条件(8)を満たさないが，これは入力文字列 x に現れなかった文字図形を，x の解析に無関係だから，Σ' に含めていないためであり，問題とはならない．

以上のパターン生成確率文法 $G^*(x)$により，文字図形列入力($G^*(x)$ の文)$x=x_1\cdot\cdots\cdot x_n$ を生成する木の中で，生起確率が最大の構文木 T^* を構文解析アルゴリズムによって求めることができる．T^* の leaf をすべて刈り取った木を $\overline{T^*}$ とすると，$\overline{T^*}$ が x を生成した G の構文木であると推定される．また，$\overline{T^*}$ の leaf 列(front) $c_1\cdot\cdots\cdot c_n$ を T^* が生成する文と呼ぶ．

$$\max_T P_r(T|x_1\cdot\cdots\cdot x_n)=P_r(T^*)$$

$$P_r(T^*)=P_r(\overline{T}^*)\cdot\prod_{i=1}^{n}P_r(x_i|c_i(\overline{T}^*))$$

$$\prod_{i=1}^{n}P_r(x_i|c_i(\overline{T}^*))=P_r(x|c_1\cdot\cdots\cdot c_n(\overline{T}^*))$$

だからである．

(3)式を最大にする文 $c_1\cdot\cdots\cdot c_n$ を $G^*(x)$から求めたい場合には，認定問題とは少し趣の異なる問題を解くことになる．これは，$x=x_1\cdot\cdots\cdot x_n$ を

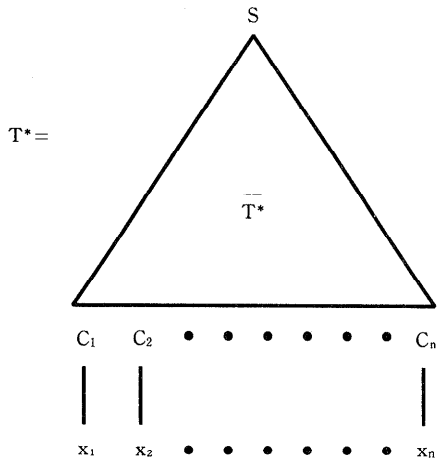


図-1 パターン生成文法における構文木

生成する $G^*(x)$ の構文木の中で，同一の文 $s=c_1\cdots\cdots c_n$ を生成する構文木の生起確率の総和が最大となる s を求める特題である．この問題を解くには，指数のオーダの時間と記憶領域を要するアルゴリズムしか発見されておらず，また，**1.** で述べたように，文字/音声認識がヒューマンインタフェースの一環として行われることが多いので，$x-x_1\cdots\cdots x_n$ を生成する最大の構文木を求めることのほうが大切であると思われる．

音声認識の場合に文字認識と異なる点は，音声認識の場合には Σ が単語(または文節)の全体集合であること，入力図形列 x が音響パラメタ時系列であることであるが，処理の本質は異なるものではない．

文字認識の場合，定められた桝目に，楷書で丁寧に文字図形を書く場合には，直前の文字図形の形状が直後の文字図形の形状に影響することは少ないが，速筆で書く場合には直前の文字図形の最終ストロークの方向と終点が直後の文字図形の開始点や開始ストロークの方向に影響する．このような影響の度合は，文字図形を正規化することにより少なくする方法がとられるが，これが困難な場合には前後の文字図形に影響し合う開始点と終点の位置，開始ストロークと最終ストロークの方向の情報を加味した文法範疇で確率文法を構成する必要がある．音声認識では調音結合がこのことに対応する．

4. 最近のトピックスと結論

文字/音声認識で確率文法が必要とされる理由と自然言語処理においても事例データを用いた構文解析の曖昧さの解消には有効な手段の一つであることをまず説明した．次に，確率文脈自由文法の二種類の定義(標準形PCFG, Conditional PCFG)およびConditional PCFGから標準形への等価変換，認定問題と構文解析法とその能率，確率パラメタの推定法，確率文法のパターン認識における応用を述べ，特に，自然言語処理における最近の傾向である事例データの言語処理への利用法に関係する，確率パラメタの推定法とその問題点をやや詳しく述べた．

自然言語における文の生成機構，すなわち文法は統語範疇(大雑把には品詞と解釈してよい)の間の文脈自由型の生成規則を基軸とし，(統語範疇の) 句と句の間の係り受けを支配する語 (主辞) の共起関係を副軸として組み立てられる．共起関係を文脈自由型の生成規則に取り入れることは，むやみに行えば，生成規則の数が膨大となり，このことは確率文法のパラメタ推定にとてつもない数の事例データを要することになり現実的ではない．

最近，共起関係が記述できる自然言語の確率文法として，確率木接合文法(Probablistic Tree Adjoining Grammar)の一変種であるProbabilistic Lexicalized Context-Free Grammarが提案されているが，文法規則が膨大になる点で問題はなんら片付いていない[5),8),9)]．筆者は，統語範疇にシソーラス(語の上位下位関係)を導入した確率文脈自由文法で語の共起関係を取り入れることを考えているが，そのことについては別の機会にゆずる．

参 考 文 献

1) Baum, L. E. et al.: A Maximalization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains, Annals of Mathematical Statistics, 41, No. 1 (1970).
2) Baum, L. E.: An Inequality and Associated Maximization Tcchnique in Statistical Estimation of Probabilistic Functions of a Markov Process, Inequalities, 3 (1972).
3) Fu, K. S.: Stochastic Language for Picturc

Analysis, Computer Graphics and Image Processing, 2 (1973).
4) Tousant, G. T.: The Use of Context in Pattern Recognition, Pattern Recognition, 10 (1977).
5) Joshi, A. et al.: Tree Adjunct Grammars, Journal of Computer and System Sciences, 10 (1975).
6) 日高，長田：日本語の文脈情報を用いた文字認識，電子通信学会論文誌，J67-D, No. 4 (1984).
7) 永井，日高：日本語における単語の造語モデルとその評価，情報処理，Vol. 34, No. 9 (1993).
8) Schbes, Y.: Stochastic Tree-Adjoining Grammars, Proc. of COLING, 1 (1992).
9) Resnik, P.: Probabilistic Tree-Adjoining Grammar as a Framework for Statictical Natural Language Processing, Proc. of COLING, 1 (1992).
10) 日高：確率文脈自由文法におけるパラメタの最尤推定法，電子情報通信学技法，NCL93-57 (1993).

付録 Conditional PCFGの標準形PCFGへの等価変換アルゴリズム

Conditional PCFG $G=(\Sigma, V, P, S, p)$ を標準のPCFG $G'=(\Sigma, V', P', S', p')$ に変換するには，次のようにすればよい．ただし，簡単のため，開始記号Sを左辺とするGの書換え規則はただ一つで，これを $\delta_0 \in P$ とする．また，書換え規則 $\delta \in P$ の左辺 $(\in V)$ と右辺 $(\in(\Sigma \cup V)^*)$ をそれぞれ δ_L, δ_R で表し，記号列 $\alpha \in (\Sigma \cup V)^*$ において，α に現れるすべての非終端記号 $X \in V$ を新たに導入された非終端記号 X_δ に置換した記号列を $\alpha(\delta)$ で表す．

【標準形への変換】

$$V' \overset{\text{def}}{=} \{X_\delta | X \in V, \delta \leftarrow P\}$$

$$P' \overset{\text{def}}{=} \{\delta_L(\delta') \rightarrow \delta_R(\delta) | \delta, \delta' \in P\}$$

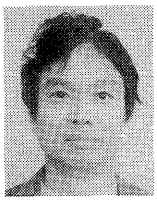
$$p' ; p'(\delta_L(\delta') \rightarrow \delta_R(\delta)) \overset{\text{def}}{=} p(\delta, \delta')$$

$$S' \overset{\text{def}}{=} S_{\delta_0}$$

□

上記の変換において，記号 $X_\delta \in V'$ は書換え規則 $\delta \in P$ を適用して出現した非終端記号 X であることを表示する非終端記号であり，$\delta_L(\delta') \rightarrow \delta_R(\delta)$ は書換え規則 δ' を適用して出現した非終端記号 δ_L に対して書換え規則 δ を適用することを表示する書換え規則になっている．したがって，G と G' が等価であることは自明であろう．

(平成6年6月13日受付)



日高 達

昭和14年生．昭和40年九州大学工学部電子工学科卒業．昭和42年同大学院工学研究科電子工学専攻修士課程修了．昭和44年同大学院工学研究科電子工学専攻博士課程中退．同年九州大学工学部助手，昭和48年同講師，昭和55年同助教授，昭和63年同教授，現在に至る．工学博士．形式言語の方程式論，自然言語処理，手書き文字認識の研究に従事．電子情報通信学会，人工知能学会各会員．