

# ストローク単位の確率文脈自由文法を用いたオンライン手書き数式認識

山本 遼<sup>†</sup> 酒向 慎司<sup>†</sup> 西本 卓也<sup>†</sup> 嵯峨山茂樹<sup>†</sup>

<sup>†</sup> 東京大学 大学院情報理工学系研究科

〒 113-8656 東京都文京区本郷 7-3-1

E-mail: †{yamaryo,sako,nishi,sagayama}@hil.t.u-tokyo.ac.jp

あらまし 本論文ではオンライン手書き数式認識のための新しい手法を提案する。我々は、手書き数式を、ストローク列を生成する確率文脈自由文法としてモデル化し、数式の文法構造と2次元構造、シンボル内の構造を全て1つの文法で統一的に記述する手法を提案する。この手法によりシンボル認識・構造認識・数式文法解析といった独立の認識ステップを経ずに、これらの情報を同時に利用して数式認識を行うことができ、頑健な数式認識が期待できる。また数式の2次元構造モデルとして、さまざまな形状の非終端記号を統一的に扱うために、隠れ筆記領域という概念を提案し、これを用いて「シンボル形状に依存しない位置」をモデル化する。このモデルはデータからのパラメタ学習が可能であるため、十分な学習データが存在すれば頑健なモデルとなる期待がある。同一筆者が入力した、学术论文における中程度複雑さの8種80数式による評価実験の結果、数式のシンボル・構造の同時推定によってシンボル分割・認識精度が向上することが確かめられた。隠れ筆記領域は学習データが非常に限定されていた中で、シンボルのベースライン正解率約95%が得られた。

キーワード 数式認識 文字認識 オンライン 手書き 確率モデル 確率文脈自由文法

## Stroke-Based Stochastic Context-Free Grammar for On-line Handwritten Mathematical Expression Recognition

Ryo YAMAMOTO<sup>†</sup>, Shinji SAKO<sup>†</sup>, Takuya NISHIMOTO<sup>†</sup>, and Shigeki SAGAYAMA<sup>†</sup>

<sup>†</sup> Graduate School of Information Science and Technology, The University of Tokyo

7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-8656 Japan

E-mail: †{yamaryo,sako,nishi,sagayama}@hil.t.u-tokyo.ac.jp

**Abstract** In this paper, we propose a new framework for on-line handwriting mathematical expression recognition. In this approach, we see handwriting mathematical expression as a stroke generation process with grammar. We do not solve 3 major problem in mathematical expression recognition one by one: symbol segmentation/recognition, 2-D structure analysis, and math syntax analysis, rather we estimate the most likely expression candidate derived from one stroke-based stochastic context-free grammar. Experiments of on-line handwriting mathematical expression recognition with 8 math expressions extracted from IEEE transactions, each of them written 10 times by one writer revealed that symbol segmentation error rate and symbol recognition error rate decline with simultaneous estimation of symbol, 2-D structure, and syntax of mathematical expressions.

**Key words** Mathematical Expression Recognition, Character Recognition, On-line, Handwriting, Stochastic Model, Stochastic Context-Free Grammar

### 1. はじめに

現在数式のコンピュータへの入力には $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ やC言語などの専用言語による入力と、エディタによる入力主流である。しかしこれらの入力方法は訓練が必要であったり操作が煩雑であるなどの面で、使いやすいとはいえない。ペンタブレット等に

よる直感的な数式の入力が実現されれば、論文などへの数式入力、数式検索や数値計算ソフトの入力インタフェースなど教育分野において幅広い応用の可能性がある。

数式認識問題はかつてよりシンボル認識と構造認識の2つの問題からなると捉えられてきた。シンボル認識はさらに、シンボル分割と(狭義の)シンボル認識の2つの問題として捉えら

れ、構造認識には2次元構造の認識と数式文法による解析の2つの側面がある。このように数式認識問題はシンボル認識と構造認識の2つの問題、さらに細かくはシンボル分割・認識問題、2次元構造認識問題、数式文法解析問題の4つの問題からなる。

現在オンライン手書き数式認識の研究は、主にオフライン数式画像認識の手法を応用して行われている。オフライン数式画像認識では、シンボルの分割・認識における高い精度が期待できるため、認識されたシンボル間の構造認識についての研究例が多い[6]。数式の2次元構造認識には、グラフ文法などの文法を用いる手法[1]や、 $xy$ 方向の逐次的な分割ルーチンにより構造を決定[2]するなどルールを細かく記述したルールベースの手法等が用いられてきた。

数式文法解析は主に構造認識の結果からの数式意味抽出として研究された。グラフ文法で記述された2次元構造の数式文法への木構造変換[3]や、2次元木構造から $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$ のコードを生成しコードの意味解析を行うテキストベースの手法[4]などが提案されている。またルールを詳細に記したルールベースのシステムでは、システム内部に暗黙的に数式文法が存在していることが多い。

オンライン手書き数式認識にも同様の構造認識が用いられており、閾値によって隣接シンボル間の構造を決定[3]したり、縦横方向の逐次的分割により構造を決定[5]する手法などがオンラインにも応用されている。

現在提案されている数式認識手法のほとんどは、最初に数式内のシンボルを認識し、その後シンボル間の位置関係から数式構造を認識するというように、シンボル認識と構造認識の2ステップからなる問題として数式認識を捉えている。しかし我々が数式内のシンボルを認識する際、そのシンボル形状のみではなく全体の構造や数式文法の情報をも利用し、シンボル形状を含めた総合的な判断を行っていると考えられる。そこで本研究では数式認識をシンボルと構造の同時推定問題と捉え、ストローク尤度と構造尤度の積を最大化する数式仮説の探索問題として定式化を行う。この探索問題はストローク尤度と構造尤度を適用確率とする確率文脈自由文法による最尤導出を求めるとして、CYKアルゴリズムにより効率的に解くことができる。

また数式の構造を評価する構造モデルとして、数式の背後に存在するシンボルに依存しない「マス目」をモデル化した隠れ筆記領域によるモデルを提案する。各ストロークが隠れ筆記領域のどの位置に書かれるかをあらかじめ統計的に学習し、認識時はストローク仮説ごとに隠れ筆記領域の尤度分布を推定し、この特徴量をもとにシンボルによらない構造評価を行う。データからシンボルの配置情報を学習するだけでなく、カンマやドット、アクセントなどの特殊な書かれ方をする文字も全て含めて統一的に構造評価が行え、文法規則の簡素化につながる。

以下2.にて提案手法について述べ、3.において評価実験を行う。

## 2. ストローク単位の確率文脈自由文法を用いたオンライン手書き数式認識

### 2.1 問題意識と解決の枠組み

数式内のシンボルの認識と構造の認識は密接に結びついている。例えば図2.1(左上)、図2.1(右上)の例はシンボルの形状が同じでも、文脈によってシンボルの認識は変わりうることを示している。このように数式内のシンボルは人間でさえ、単独で認識できるとは限らない。人間はこの曖昧性を、数式の構造や文法を用いて補っていると考えられる。このように数式認識においてもシンボルの認識と構造の認識を同時に行うことで認識精度が向上する期待がある。これはシンボルの尤度と構造の尤度の積を最大化する数式仮説の探索問題と捉えることにより定式化が可能となる。

さらに、数式のシンボルの境界も、シンボル情報のみから一意に認識できるものではない。図2.1(下)の例ではそれぞれ数式文法の制約、数式構造情報により“ $x$ ”や“括弧”などとシンボルの境界が変わる。シンボル分割、シンボル認識、構造認識を全て総合的に判断することにより、シンボル境界の決定が頑健に行えると考えられる。そこで、構造を考える最小単位をシンボルではなくストロークとし、ストロークのなす構造としてシンボルを捉える発想が生まれる。すなわストロークが構造を持ってシンボルとなり、シンボルが構造を持って数式になるというように、シンボルと数式構造を同じ「構造」と捉えて統一的に扱う。この手法は、シンボル分割・シンボル認識・構造認識の3つのステップを全て統一的に扱うことができる。

また既存の数式認識手法の多くは、数式の2次元構造と数式文法が分離しているが、これは構造認識の結果から必ずしも数式文法への変換が保障されない点に問題があるといえる。我々は数式に内在する文法をベースに構造文法を作成することにより、2次元構造と数式文法構造の同時が可能であると考えた。

ここまで

- ストロークを単位とする
- 数式文法をベースに作成する
- シンボルの構造と数式の構造を統一的に扱う文法

により、シンボル分割・シンボル認識・構造認識・文法解析の4つの問題を全て同時に解くという問題意識を述べた。この枠組みの定式化について次項にて述べる。

### 2.2 定式化

数式文法から生成可能な数式仮説集合を  $E_X$  とし、手書き数式を  $H$  とすると、数式認識とは

$$X_0 = \arg \max_{X \in E_X} P(X|H) \quad (1)$$

$$= \arg \max_{X \in E_X} P(H|X)P(X) \quad (2)$$

$$\simeq \arg \max_{X \in E_X} P(H|X) \quad (3)$$

となる  $X_0$  を推定する問題である。ここで  $P(H|X)$  は数式仮説  $X$  から手書き数式  $H$  が生成される尤度、 $P(X)$  は数式仮説  $X$  の出現確率である。今数式の出現確率は一律であると仮定した。

数式はストロークがそれぞれの形状と位置をもって配置され

$$\begin{array}{l}
 2 < x+d \\
 2 < (2x+d)^2
 \end{array}
 \quad
 P(x|y, z)$$

$$(a+2)(a-2) \quad (a+2)(a^2-2)$$

図 1 (左上) シンボル形状が同じでも数式文法によって解釈が変わる例。2 画目のストロークは上では“c”，下では“括弧”と解釈される。(右上) シンボル形状と数式構造を同時に評価して認識すべき数式例。“ $P(x|y, z)$ ”と“ $P(x_1y, z)$ ”のどちらかを認識する際、人間は、“ $x$ ”と“ $y$ ”の間の記号がどれくらい“縦棒”らしいか、もしくはどれくらい“1”らしいかというシンボル形状の評価と、その位置がどれくらい“ $x$ ”の横らしいかもしくは右下らしいかという構造の評価を同時に行い、総合的に判断していると考えられる。(下) 構造の変化がストローク認識に影響を与える例。中央は上では括弧である可能性があるが下では  $x$  と解釈するのが自然である。

たものである。そして適切な形状を持つストロークが複数、適切な位置関係をもって配置されているとき、それらはシンボルとみなすことができる。そして適切なクラスに属するシンボル同士が、また適切な位置関係に配置されているとき、それらは「項」や「式」、「分数」などの数式文法の要素となる。そしてこれらの適切な要素同士が再び適切な位置関係に配置されてるとき、それらは全体として数式を構成する。この数式の構成は、表 1 のような確率文脈自由文法による導出として表せる。

ここで用いる確率文脈自由文法  $G$  の生成規則は、非終端記号である数式部品の集合  $V_N$ 、終端記号である手書きストローク (今後単に「ストローク」は「 $x$ 」の 1 ストローク目) などのシンボリックな情報を表し、「手書きストローク」はペンタブレット等によりサンプリングされたストロークの点列 (形状) 情報を表すこととする。) の集合  $V_T$ 、開始記号「式」である  $S$ 、構造条件付き非終端記号生成規則の集合  $P$ 、終端記号生成規則の集合  $Q$  からなる。構造条件付き生成規則は  $p = \langle A \rightarrow BC, s \rangle$  の形をしている。 $A, B, C \in V_N$  であり  $s$  は「右」「右上」等の構造を表す。構造  $s$  の構造モデルは  $P(B, C|A, s)$  を与える。構造モデルの詳細は 2.4 にて述べる。

終端記号生成規則は、 $q = \langle t \rightarrow \alpha \rangle$  の形をしている。 $t \in V_N$  はストロークであり、 $\alpha \in V_T$  は手書きストロークである。ここで  $\alpha$  が  $t$  から  $q$  により直接導出される尤度は、 $P(\alpha|t)$  であり、これを  $t$  ごとにストロークモデルによりモデル化する。ストロークモデルは入力ストロークのモデル尤度を返す関数で、文字認識のための文字モデルが利用できる。

今手書き数式  $H$  の文法  $G$  による導出  $X = \{p_1, p_2, \dots, p_N, q_1, q_2, \dots, q_M\}$  において、 $n$  番目に適用された非終端記号生成規則を  $p_n = \langle A_n \rightarrow B_n C_n, s_n \rangle$ 、 $m$  番目に適用された終端記号生成規則を  $q_m = \langle t_m \rightarrow \alpha_m \rangle$  と表す。 $N, M$  はそれぞれ導出の非終端記号生成規則、終端記号生成規則の数である。このとき式 1 は

$$X_0 = \arg \max_{X \in E_X} \prod_{n=1}^N P(p_n) \prod_{m=1}^M P(q_m)$$

と表される。すなわち数式認識問題は、数式文法により生成される数式仮説内、ストローク尤度と構造尤度の積を最大化する数式仮説の探索問題として定式化できる。

### 2.3 CYK アルゴリズムによる解探索

確率文脈自由文法において最尤数式仮説の探索は CYK アルゴリズムにより効率的に行うことができる。図 2.3 の例についてこのアルゴリズムを述べる。

以下に手順を示す。

(1) 各手書きストロークについて、全ストロークモデルそれぞれから生成された尤度 (ストローク尤度) を計算し (文字認識の尤度計算とほとんど同じ)、CYK マトリクス対角成分  $\text{stroke\#1} - \text{stroke\#4}$  にストロークと尤度を記入。例えば 1 ストローク目の形状は“2”や“z”に近いので、それらの候補の尤度が高くなる。実際には候補の上位  $N$  位までを候補として書き込む。

(2) 第  $i$  ストロークと第  $i+1$  ストロークに対して適用可能な規則全てを適用し、ストローク尤度と構造尤度の積をマトリクスの  $(i, i+1)$  成分に記入する。例えば  $\text{stroke\#1}$  と  $\text{stroke\#2}$  に規則を適用すると“ $2a$ ”や“ $2^a$ ”等の導出が存在する。“ $2a$ ”の構造尤度は第 1 ストロークと第 2 ストロークの位置関係が「横」である構造尤度で、“ $2^a$ ”は「右上」である構造尤度である。これらを次項に述べる構造モデルから求め、それを各生成規則の適用確率とする。マトリクスに書き込まれる尤度は、第 1・2 ストロークのストローク尤度と構造尤度の積である。

(3) 次にマトリクスの  $(i, i+2)$  成分を記入する。第  $(i, i+2)$  成分には、第  $(i, i)$  成分と第  $(i+1, i+2)$  成分の組合せに対して適用可能な規則を適用した場合と、第  $(i, i+1)$  成分と第  $(i+2, i+2)$  成分の組合せに対して可能な規則を適用した場合の 2 通りがある。例えば  $\text{stroke\#1} - 3$  を記入する場合、 $\text{stroke\#1}$  と  $\text{stroke\#2} - 3$  に規則を適用する場合と  $\text{stroke\#1} - 2$  と  $\text{stroke\#3}$  に規則を適用する場合を考える。前者の例としては“ $2a^2$ ”等が、後者の例としては“ $2az$ ”などがある。これらの構造尤度を求め、規則を適用した各数式部品の尤度、構造尤度の積をマトリクスに記入する。

(4) 以下マトリクスの  $(i, i+j)$  成分を記入する時は、 $k = 0, \dots, k-1$  として  $(i, i+k)$  成分と  $(i+k+1, j)$  成分のすべての組合せそれぞれに対し、適用可能な規則を適用した結果を書き込む。

(5) 最終的に  $\text{stroke\#1} - 4$  の部分に書き込まれた候補のうち、開始記号である「式」でありかつ尤度が最も高い候補が認識結果となる。

### 2.4 隠れ筆記領域を用いた構造モデル

数式部品間の構造の評価には、多くの研究で外接矩形を用いた評価が用いられている (図 3(左) 等)。しかし外接矩形は各シンボルに依存するため、外接矩形のみから構造を評価するのは難しい (図 3(右))。我々はこの外接矩形の背後に、シンボルによらない矩形領域の存在を仮定する構造モデルを用い、頑健な構造評価を目指す。以下にこのモデルの概要を述べる。

数式が書かれる際すべての手書きストロークの背後には、その数式の文法構造から確率的に定まる、ある「整ったマス目」

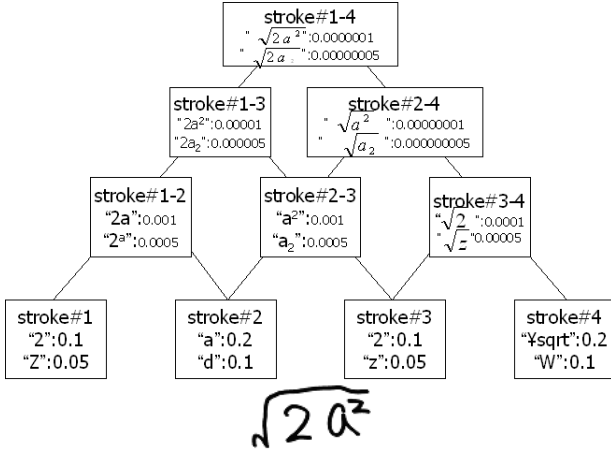


図2 CYK アルゴリズムによる最尤数式仮説探索の例．この手書き数式はストロークが”2”，”a”，”2”，”根号”の順序で書かれたとする．

が存在すると仮定する ( 図??(左上) ) ．そしてその「整ったマス目」に対してそれぞれの手書きストロークが，そのストロークの種類に依存した揺らぎを持って書かれるとする ( 図??(中央) ) ．この手書き数式の背後に存在する「整ったマス目」を隠れ筆記領域と呼ぶことにする ．

上の内容を定式化する．数式部品  $A$  の隠れ筆記領域を  $h_A$  とし， 図??(右上) のように  $h_A = \{h_A^c, h_A^s, h_A^b, h_A^e\}$  の4座標で表す． 図??(下) のように生成規則  $\langle A \rightarrow B C, s \rangle$  の構造条件  $s$  ごとに，  $h_A, h_B, h_C$  の満たす制約条件を設定する．制約条件は  $h_A, h_B, h_C$  の各座標と，構造固有の確率変数  $v_s$  からなる連立方程式  $W_s$  で記述される．数式部品  $A$  から  $B, C$  が生成規則  $\langle A \rightarrow B C, s \rangle$  によって生成される確率は，

$$P(B, C | A, s) = \int P(h_B, h_C, v_s | h_A, W_s) P(v_s) dv_s \quad (4)$$

$$= \int \delta(W_s) P(v_s) dv_s \quad (5)$$

となる．  $\delta(W_s)$  は  $h_A, h_B, h_C, v_s$  の関数で，連立方程式  $W_s$  が成立する時 1 をとり，それ以外で 0 となる関数である ．

隠れ筆記領域  $h_A$  からのストローク位置  $r_\alpha$  の決定は， 隠れ筆記領域  $h_A = \{h_A^c, h_A^s, h_A^b, h_A^e\}$  と手書きストローク外接矩形  $r_\alpha = \{r_\alpha^c, r_\alpha^s, r_\alpha^b, r_\alpha^e\}$  のずれ  $r_\alpha - h_A = d_A = \{d_A^c, d_A^s, d_A^b, d_A^e\}$  の確率分布  $P(d_A)$  をストロークごとに設定し，生成確率を

$$P(\alpha | A) = \delta(r_\alpha - h_A - d_A) \quad (6)$$

とする ．

観測ストローク外接矩形列を  $\{r_1, r_2, \dots, r_M\}$  ，これに対する導出を  $X = \{p_1, p_2, \dots, p_N, q_1, q_2, \dots, q_M\}$  ( $p_n = \langle A_n \rightarrow B_n C_n, s_n \rangle$  ,  $q_m = \langle t_m \rightarrow \alpha_m \rangle$ ) とする ．この導出の尤度は

$$\begin{aligned} \prod_{n=1}^N P(p_n) &= \prod_{n=1}^N P(B_n, C_n | A_n, v_s) \prod_{m=1}^M P(\alpha_m | t_m) \\ &= \int \prod_{n=1}^N P(h_{B_n}, h_{C_n} | h_{A_n}, v_s) \prod_{m=1}^M P(r_m | h_m) dh \end{aligned}$$

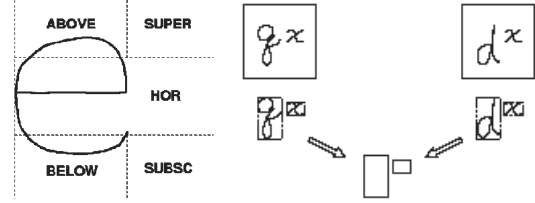


図3 (左)Zanibbi らの用いたシンボル間位置関係の分類．隣接文字の重心が HOR の領域に存在すれば横に並んでいと認識，SUPER の領域であれば指数と解釈． [3] より引用 ．(右) 外接矩形による位置関係評価の限界示す例 ．  $q^x$  と  $d^x$  のシンボル間の構造は異なるが，外接矩形は等しい ．[6] より引用 ．

となり，これらは式5，式6を用いて解くことができる ．

各構造のパラメタ  $P(v_s)$  と各ストロークのパラメタ  $P(d_A)$  それぞれの確率分布の学習は，学習データにおける最尤隠れ筆記領域の尤度を最大化するように以下の反復推定を行う ．まずパラメタに初期値  $P(v_s)$  と  $P(d_A)$  正解数式構造が既知である各学習データから最も尤度の高い隠れ筆記領域を推定し，各シンボルごとにその隠れ筆記領域と外接矩形のずれの分布を再推定し，各構造ごとに確率変数パラメタを再推定する ．そして最推定されたモデルパラメタを用いて再び最尤隠れ筆記領域推定とそれによるパラメタ再推定を行うという反復推定を用いた ．

隠れ筆記領域による構造モデルを評価するため，数式構造認識実験を行った ．図5の#4-#8を含む7本の数式について1筆者が20~40回ずつ書いた手書き数式254を構造モデルの学習データとし，別個に図5全8数式を同じ筆者が10回ずつ書いた手書き数式80を評価データとした ．その結果，ベースラインの誤り率  $E_{str}$  は学習データに含まれる数式(#4-#8)で2.53%，含まれない数式(#1-#3)で5.07%，トータルで3.12%であった ．また各データにおいて推定された最尤隠れ筆記領域は図6のようになった ．

学習データが7種類の数式のみと限られていることを考えると，効率的に構造モデルの学習を行うことができたといえる ．また構造の認識誤りは，(1) 数式が傾いていたため式の一部を右上，右下としてしまう誤認識，(2) 学習データと評価データで異なる文脈・サイズで出現するシンボルの誤認識などが存在した ．(1) は傾きを考慮したモデルの修正により，(2) は学習データ量を増やすことによりそれぞれ改善が可能と考えられる ．

### 3. 評価

本手法の有効性を検証するため，以下の4種類の数式文法を用いた場合について数式認識実験を行い，その性能を比較・評価した ．

(1) 数式構造を評価しない文法を用いた場合 ．表1の18番目以降の生成規則と，表2の「不特定」構造規則を用いて認識 ．不特定構造の構造尤度は，シンボル間の位置関係によらず常に一定の値  $\gamma$  である ．この文法のみを用いてシンボルを構成することで，数式構造情報を利用しない場合のシンボル認識精度を調べた ．

(2) 制約の弱い文法を用いた場合 ．表1の括弧に関する

表 1 簡易な数式文法．数式としての構造制約に，括弧を閉じる制約を付加したもの． $\text{L}^{\text{A}}\text{T}_{\text{E}}\text{X}$  で記述可能なほとんどの数式のうち，括弧が閉じられているものを生成可能．備考\*:規則が不適切に繰り返されないように実際はより細かい場合分けを行っていることを示す (例えば 5 番目の規則が繰り返し適用されて特殊関数の上にいくつも式が生成されることはない)，\*\*:書き順が逆である規則も含む，\*\*\*:実際は括弧の種類による場合分けを行っていることを示す

No.	生成規則	構造	備考
1	式 文字		
2	式 式 文字	右	
3	文字 文字 式	右上	*
4	文字 文字 式	右下	*
5	特殊関数 特殊関数 式	上	*
6	特殊関数 特殊関数 式	下	*
7	分母付き分数線 分数線 式	分数下	**
8	分子付き分数線 分数線 式	分数上	**
9	文字 分母付き分数線 式	分数上	**
10	文字 分子付き分数線 式	分数下	**
11	文字 根号 式	内部	**
12	文字 式アクセント 式	アクセント位置	** **
13	文字 文字アクセント 文字	アクセント位置	** **
14	右括弧付き式 式 右括弧	右	** ** **
15	左括弧付き式 式 左括弧	左	** ** **
16	文字 右括弧付き式 左括弧	左	** ** **
17	文字 左括弧付き式 右括弧	右	** ** **
18	文字 $a b c \dots$		
19	特殊関数 $\lim \sum \max \dots$		
20	左括弧 $(\{[\{ \dots$		
21	右括弧 $\} \}\} \dots$		
22	f $f_1 f_2$	同一シンボル	
23	x $x_1 x_2$	同一シンボル	
24	:		
25	a 手書きストローク		
26	$f_1$ 手書きストローク		
27	分数線 手書きストローク		
28	:		

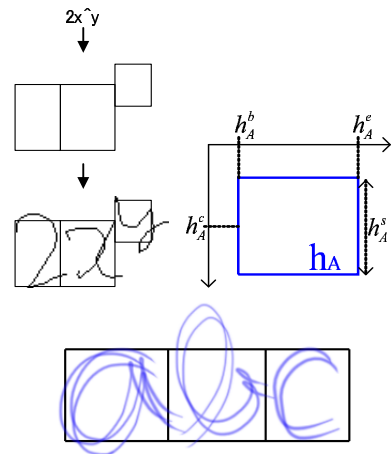
文法 (7 番から 10 番) を用いず，括弧を他の文字と同様に扱う文法を用いた．

(3) 括弧の制約を持つ文法を用いた場合．表 1 の生成規則を用いた．

(4) より強い制約の文法を用いた場合．積分は  $d$  で終わる，項と項の間に演算子が現れるなどの制約を追加した文法を用いた．

実験 1 と実験 2, 3, 4 を比較することで，数式構造の評価がシンボル分割・認識精度にどのように影響するかを調べた．

ストロークモデルには正規化座標と速度ベクトル  $(x, y, v_x, v_y)$  の時系列を特徴量とした 15 状態の Left-to-Right HMM を用いた．各状態からの出力確率は対角共分散単一正規分布とした．実験の学習データと評価データは 2.4 の隠れ筆記領域の評価実験と同じものを用いた．シンボル分割・認識の精度指標として



位置関係	図示	条件定式化
横		$h_b^c = h_a^c = h_a^s, h_b^s = h_a^c = h_a^s,$ $h_b^c = h_a^c, h_b^s = h_a^c, h_b^s = h_a^c$
右上		$h_b^c = h_a^c, h_b^s = h_a^s, h_b^c = h_a^s,$ $h_c^c = h_a^c, h_b^s = h_a^c,$ $h_c^c = h_b^c - 1/2 h_b^s - q_1,$ $h_c^s = 1/2 h_c^c + q_2,$
アクセント 同一シン ボル		$h_b^c = h_a^c = h_a^s, h_b^s = h_a^c = h_a^s,$ $h_b^c = h_a^c = h_a^s, h_b^s = h_a^c = h_a^s$

図 4 (左上) 隠れ筆記領域の概念図 (右上) 隠れ筆記領域を表すパラメータ (中央) 隠れ筆記領域に対するストロークの確率的にストロークが書かれるイメージ図 (下) 各構造におけるの隠れ筆記領域のみたす条件

$$\#1 \delta_{B(x)} = 2\gamma \left[ \left(\frac{\sigma}{\gamma}\right)^2 - \left(\frac{\sigma}{\gamma}\right)^2 \right] - \delta_{B(x)}(2|\sigma|)$$

$$\#2 \Gamma(l, m, n) = \sum_{k=1}^{2n} (e^{(k)}) \Gamma_n(m, R)$$

$$\#3 E_p(n) = \frac{\hat{R}}{s(2\sqrt{n} - s_{n+1})}$$

$$\#4 \frac{p(\Gamma(d))}{p(\Gamma_0(d))} = \int d(s, f) \left[ \frac{p(\Gamma(s))}{p(\Gamma_0(s))} \right] p(s, f | \Gamma_0, d)$$

$$\#5 d(x) = \left\{ \int dn^s S[n(x)] - n^s \int dn^b B(n^s, n^b) S(n^s) \right\} + \eta(x)$$

$$\#6 p(d|G) \propto \int ds e^{-(1/2)(d-Rs)N_A^{-1}(d-Rs)} e^{-(r/2)s^2 C^{-1}s}$$

$$\#7 \sigma_k(S^{(1)}, S^{(2)}) = \frac{1}{2k+1} \sum_{l=0}^k \|l\|_m (S^{(1)} + S^{(2)})^l$$

$$\#8 W_p(\gamma_p) = 2 \int_0^{\gamma_{max}} \xi(\sqrt{\gamma_p^2 + y^2}) dy$$

図 5 評価に用いた手書き各数式のデータ例．左側は数式番号．#4 から #8 は学習データと数式が重なっている．

シンボル分割誤り率  $E_{seg}$  とシンボル認識誤り率  $E_{sym}$  を用いた．これは正解シンボルのうちどれだけが誤りであることを表す量である．正解数式のシンボル分割を  $A = \{a_1 = (1, i_1), a_2 =$



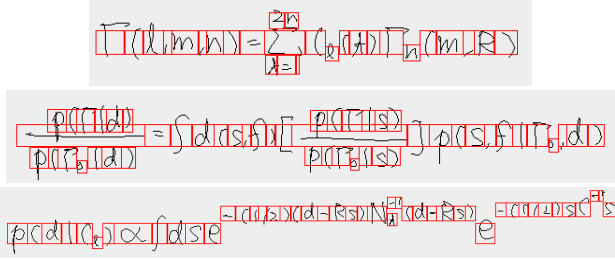


図 6 推定された最尤隠れ筆記領域．カンマなど特殊な形状のシンボルも統一的に扱うことができています．

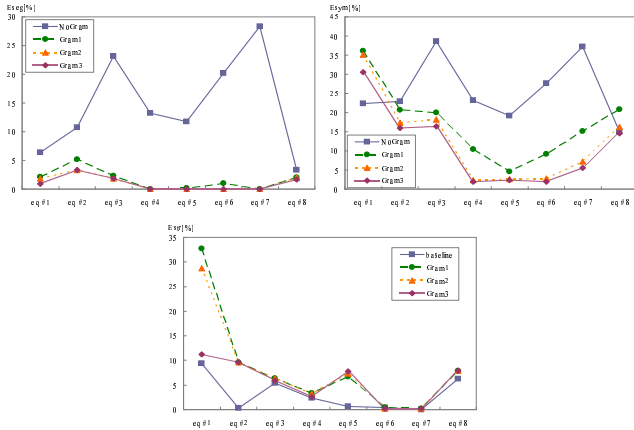


図 7 (左上)(右上) 各文法を用いたときのシンボル分割誤り率とシンボル認識誤り率．グラフでは実験 1 から 4 をそれぞれ No-Gram, Gram1, Gram2, Gram3 で表した．(下) 各文法を用いたときのベースライン誤り率．シンボル認識に正解を与えた場合を baseline で示した．

表 2 評価実験 1 において用いる生成規則

No.	生成規則	構造	備考
1	シンボル シンボル シンボル	不特定	
2	シンボル 文字   特殊関数   左括弧   右括弧		

$(i_1+1, i_2), \dots, a_n = (i_{n-1}+1, i_n)$  , 認識結果のシンボル分割を  $B = \{b_1 = (j_1, j_2), b_2 = (j_2+1, j_3), \dots, b_m = (j_{m-1}+1, j_m)\}$  とする．ここで  $a_k = (i, j)$  は, 第  $i$  ストロークから第  $j$  ストロークまでが第  $k$  シンボルを構成することを示す．また  $n, m$  は正解数式と認識結果数式それぞれに含まれるシンボル数．また正解数式のシンボル列を  $S = \{s_1, s_2, \dots, s_n\}$  , 認識結果数式のシンボル列を  $T = \{t_1, t_2, \dots, t_m\}$  とするとき,

$$E_{\text{seg}} = \frac{N(\{(i, j) | a_k = (i, j), b_l = (i, j)\})}{n} \quad (7)$$

$$E_{\text{sym}} = \frac{N(\{(i, j) | a_k = (i, j), b_l = (i, j), s_k = t_l\})}{n} \quad (8)$$

と定義する．( $N$  は集合の要素数を表す．) また不特定構造文法の出力確率  $\gamma$  は,  $10^{-3}$  から  $10^{-10}$  まで離散的に変化させて実験を行い最もシンボル誤り率の低かったものを使用した．結果は図 3.(左上), 図 3.(右上) のようになった．

数式構造を利用しない場合と比べて, 数式制約を強くすることによってシンボル分割・認識精度が向上することが確かめられた．次に文法制約によって構造認識の精度が影響を受けるかを調べるため, 各実験結果をシンボルのベースライン誤り率で

$$p(d|Cl) \propto \int ds e^{-\|R\|(d-Rs)Nt + (d-Rs)e^{\|R\|s}C + s}$$

$$p(d|C_i) \propto \int ds e^{-((1/2)(d-Rs)Nt^2 + (d-Rs)e^{-((1/2)s}C^{-1}s)}$$

$$p(d|C_i) \propto \int ds e^{-(1/2)(d-Rs)Nt^2 + (d-Rs)e^{-(1/2)s}C^{-1}s}$$

$$p(d|C_i) \propto \int ds e^{-(1/2)(d-Rs)Nt^2 + (d-Rs)e^{-(1/2)s}C^{-1}s}$$

$$p(d|C_i) \propto \int ds e^{-(1/2)(d-Rs)Nt^2 + (d-Rs)e^{-(1/2)s}C^{-1}s}$$

図 8 誤り訂正の例．上から順に同じ手書き数式の実験 1,2,3,4 の認識結果．5 番目は正解数式．実験 1 の認識結果はシンボル認識の結果を正解数式の構造に整列して表示している．

評価した．その結果が図 3. である．数式文法の強化によって構造の認識精度が上がる事が確かめられた．誤り訂正の例を図 3. に示す．構造情報, 文法情報を段階的に導入することで誤り訂正効果が段階的に大きくなっていくことがわかる．

#### 4. まとめと課題

数式内のシンボルの認識には数式構造情報の利用が有効であるという視点から, 数式認識を数式仮説の範囲内でのシンボルと構造の同時推定問題と捉え, ストローク尤度と構造尤度の積を最大化する数式仮説の探索問題として定式化を行った．ストロークを単位とした確率文脈自由文法により手書き数式のモデル化を行い, CYK アルゴリズムにより最尤仮説の探索が可能であることを示した．構造モデルとしては隠れ筆記領域によるモデル化を提案した．本手法による数式認識が可能であること, 数式構造情報の利用によりシンボル分割・認識精度が共に向上することを確認した．

今後はより多くのデータによる学習と性能評価, 数式の出現確率のモデル化による精度向上を目指す．

#### 文 献

- [1] A. Kosmala, G. Rigoll, S. Lavirotte and L. Pottier: "On-Line Handwritten Formula Recognition using Hidden Markov Models and Context Dependent Graph Grammars", Proc. Int. Conf. Document Analysis and Recognition (ICDAR), pp. 107-110 (1999).
- [2] J. Ha, R. M. Haralick and I. T. Phillips: "Understanding Mathematical Expressions from Document Images", Proc. Int. Conf. Document Analysis and Recognition (ICDAR), pp. 956-959 (1995).
- [3] R. Zanibbi, D. Blostein and J. R. Cordy: "Recognizing Mathematical Expressions Using Tree Transform", IEEE Trans. Pattern Anal. Machine Intell., **24**, 11, pp. 1-13 (2002).
- [4] K. F. Chan and D. Y. Yeung: "An Efficient Syntactic Approach to Structural Analysis of On-line Handwritten Mathematical Expressions", Pattern Recognit., **33**, pp. 375-384 (2000).
- [5] U. Garain and B. B. Chaudhuri: "Recognition of Online Handwritten Mathematical Expressions", IEEE Trans. Sys. Man Cybern. Part B: Cybern., **34**, 6, pp. 2366-2376 (2004).
- [6] K. F. Chan and D. Y. Yeung: "Mathematical Expression Recognition: A Survey", Int. J. Document Anal. Recognit., **3**, 1, pp. 3-15 (2000).