I-018

ストロークをベースとした確率自由文脈文法による手書き数式の認識

On-line Handwriting Mathematical Expression Recognition using Stroke-Based Stochastic Context-Free Grammar

山本 遼 † 山本 隼 † 西本 卓也 † 嵯峨山 茂樹 †

Ryo Yamamoto Jun Yamamoto Takuya Nishimoto Shigeki Sagayama

1. はじめに

本稿では、確率文脈自由文法による確率的なオンライン手書き数式認識の手法を提案する。

現在コンピュータへの数式入力は、TEX や matlab などの専用言語によるテキスト記述やマウスで操作する数式エディタ等が主である。しかしこれらの入力方法はトレーニングが必要であり直感的ではない。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、論文等への数式入力、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研究分野において幅広い応用の可能性がある。

2. 提案手法の概略

2.1 問題の定式化

オンライン手書き数式認識に関しては多くの既存研究がある [1]。これらの研究では数式の認識をシンボル認識と構造認識の 2 段階に分けて行う。まず入力ストローク系列をシンボル単位に分割して認識を行う。そして認識されたシンボル同士をそれぞれの位置関係を用いたルールにより結合し数式を構成する。

このようなルールベースのアプローチに対し、我々は確率的なアプローチを提案する。数式は必ず厳密な文法に沿って記述されるが、図 1 のように手書き数式はその揺らぎのため必ずしも一通りに解釈されるとは限らない。従って様々な数式仮説の中で尤度最大となるものを求める確率的な認識が本質的であると考えられる。また既存手法においてはストローク認識と構造認識は完全に分離した問題として扱われていたが、図 2 のように構造の認識とストロークの認識は切り離すことができない問題である。そこで我々は数式の認識を、ストローク尤度と構造尤度の和を最大化する数式仮説を求める問題として定式化する。

ストローク尤度とはストロークモデルから実際にその手書きストロークが生成される確率であり、構造尤度とは数式仮説から実際にそのストローク同士の位置関係が生成される確率である。ストローク尤度は既存のモデルベースの文字認識手法により尤度計算を行うことができる。構造尤度は次に述べる確率文脈自由文法 (Stochastic Context-Free Grammar:SCFG) によりモデル化が可能である。

2.2 確率文脈自由文法による数式構造のモデル化

文脈自由文法 (Context-Free Grammar:CFG) は $A \rightarrow B, AB \rightarrow C, ABC \rightarrow D$ 等の形で記述される文法で、数

図 1: 数式認識の曖昧性を示す例。$2x$ なのか 2^x なのかは人間にとっても必ずしも明らかではない。

図 2: 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。

生成規則	位置関係による適用確率
a,b,c → 文字	-
文字 → 式	-
文字 + 式 → 式	「文字の右上に式がある」尤度
式 + 式 → 式	「式同士が横に連結する」尤度
文字 + 式 → 式	「文字の右下に式がある」尤度

図 3: 数式を生成する確率文脈自由文法の例。

式のように再帰性をもつ言語の生成・解析に適する。例えば図 3 にあるような単純な数個の数式生成規則を繰り返し適用することにより、a, b, c の 3 文字からなる様々な数式 (例えば ab^{ab^c} など) が解析可能である。

確率文脈自由文法 SCFG は CFG における生成規則が適用確率をもつ文法である。例えば上から 3 番目の生成規則の適用確率は、「いかに文字に対し式の位置が右上らしいか」により定まる。このように手書き数式の構造尤度はシンボル間の位置関係により適用確率が定まる SCFG によりモデル化ができる。SCFG による構造尤度は CYK-algorithm により効率的に計算可能である。

2.3 位置特徴量 Virtual Bounding Box

シンボル間の位置関係の評価は単純な問題ではない。位置関係の特徴量として多く用いられる Bounding Box(最小囲み長方形) では図 4 のように多様なシンボル形状を統一的に扱うことはできず、文字カテゴリごとに場合分

†東京大学, The University of Tokyo

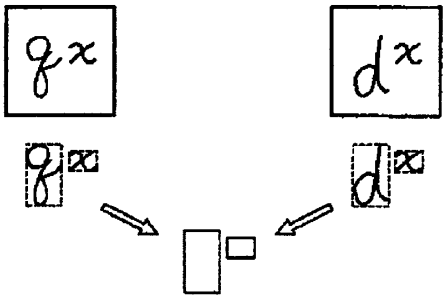

図 4: Bounding Box の限界を示す例。qx と d^x のシンボル同士の位置関係は異なるが、Bounding Box は等しい。[1] より引用。

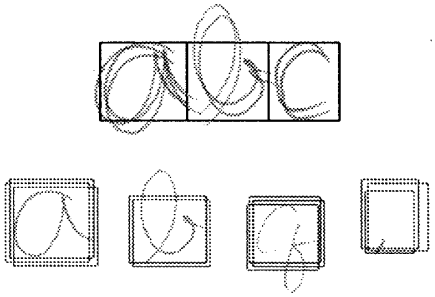

図 5: VBB の概念図。数式 abc は上の図ような長方形領域に確率的なゆらぎを持って書かれると考える。このとき a や c と比べて b は長方形領域の上方に書かれがちである。一方、実際に下の図のように書かれた文字がある場合、逆にその文字の「書かれようとした長方形領域」は尤度分布を持つ。この「書かれようとした領域」の尤度分布を Virtual Bounding Box と呼び、全てのシンボルの位置を統一的にこの特徴量で扱う。

けする必要が生じる。数式にはアルファベット以外にもアクセント、ドット、カンマ等特殊な形状のシンボルが多いため場合分けの数が多くなり、結果として SCFG の規則が複雑化してしまう。

この問題を回避するため、様々なシンボルを統一的に扱う特徴量として我々は Virtual Bounding Box(VBB) を提案する。VBB はシンボルが「どこに書かれようとしたのか」の尤度分布である (図 5)。この VBB によりアルファベット、ドット、カンマやアクセントを全て同じ特徴量で扱うことができると考えられる。以上が SCFG+VBB による確率的な数式認識手法の概略である。

3. 動作確認実験

3.1 VBB の有効性の確認

IEEE Transaction の中で中程度の複雑さを持つと思われる 8 数式 (ストローク数 30 ～ 60) 同一筆者による各 10 回ずつ計 80 の手書き数式を評価データとし、ストロークの認識が完全に行われた条件下で構造認識実験を行った。各ストロークの VBB を定めるパラメタはヒューリスティックに決定し、数式文法は図 3 のものの他に分数と平方根およびアクセントを許容した。その結果 80 数式中 63 数式の構造が完全に認識された。また誤りはサブスクリプトのミス等の局所的なもののみであり、パラメタのチューニングによりさらに精度の向上が可能であると考えられる。この実験により VBB が構造尤度計算の特徴量として有効であることが確認された。

3.2 システム全体の動作確認実験

3.2.1 ストローク尤度を手動で設定した場合

現時点では学習データやストローク尤度計算部の検討は不十分であるが、システム全体の動作を確認した。始めにストローク尤度を人手で与えた実験を行った。3.1 で用いた 8 数式各 1 回ずつ計 8 の手書き数式を評価データとし、ストロークの認識誤りの構造認識による補正を確認するため、各ストロークのストローク尤度 1 位候補を全てアルファベットの"a"、2 位を正解ストロークとし、数式の認識を行った。その結果最終的に正解ストロークが多数認識結果に含まれた。この実験により、ストローク認識誤りが構造認識において補正されうることが確認された。

3.2.2 ストローク尤度計算に HMM を用いた場合

次にストローク尤度を隠れマルコフモデル HMM により計算し、3.2.1 と同様の評価を行った。HMM の特徴量は正規化 XY 座標、正規化 XY 方向速度ベクトルの 4 次元ベクトル時系列、状態数は 10 とした。学習には 5 数式を各 40 回計 200 数式を用いた。その結果分数等の大きな構造は正しく認識できたもののシンボル認識誤りが非常に多く見られた。ストローク尤度計算精度が不十分であることが原因と考えられるが、これは学習データ量や特徴量の改善等、今後向上の可能性が十分にあると考えられる。

またこの認識実験には 1 数式につき 1 時間以上といった非常に長い計算時間が必要であった。この点についてはストローク尤度計算精度の向上により効率的な枝刈りが行える期待があり、また分数線などのはっきりとした数式の区切れを活用することで、探索空間を減らすことも可能と考えられる。

4. まとめ

確率文脈自由文法と Virtual Bounding Box による確率的な数式認識手法を提案した。実験により VBB の有効性、確率的数式認識の動作が確認できた。今後はストローク尤度計算部の精度向上が課題となる。

参考文献

[1] K.F.Chan, D.Y.Yeung: "Mathematical Expression Recognition: A Survey" *'Int. J.Document Anal.*, vol.3,no.1,pp.3–15,2000.