

## 筆記途中の手書き漢字を学習した早期認識手法

Early Recognition Method Trained on Partially Written Handwritten Kanji Characters

重杉 健士朗<sup>†</sup> 坂井 浩志郎<sup>†</sup> 中井 満<sup>†</sup>

Kenshiro Jusugi Koshiro Sakai Mitsuru Nakai

## 1. はじめに

コンピュータへの手書き入力は、直感的で自由な操作が可能であり、教育現場やクリエイティブな作業でよく用いられている。キーボード入力と比べて筆記に時間がかかるという課題があり、時間効率性が求められる環境では普及の妨げとなっているが、入力内容を予測して補完することで筆者の負担を減らすことができる。従来法としては、1文字目を認識して、これに続く文字候補を手書き風フォントで貼付ける手法 [1] がある。我々は更に早くするために1文字の筆記途中で文字を認識し、残りの筆跡を生成補完する [2]。本発表では、筆記途中の漢字データを学習し、書き始めの段階で文字を認識する早期認識手法を提案する。

## 2. データおよび識別器の構成

本研究では、JAIST IJPL のオンライン手書き文字データベース  $\gamma$  セット [3] の 60 人分の漢字 1016 文字を用いた。ある被験者 1 人が筆記したデータを図 1 に示す。筆記していないときのペンの軌跡は取り除く。ペン先の  $x$  座標,  $y$  座標, 筆圧, ペンの  $x$  座標方向への傾き,  $y$  座標方向への傾きの 5 つの特徴量を用い、手書き文字認識の識別器には RNN を用いる。中間層は 3 層にして、Gated Recurrent Unit (GRU) を用い、出力層は認識対象の 1016 文字の漢字に対応する 1016 ノードで構成する。

## 3. 筆記途中のサンプルを用いた早期認識

## 3.1 データ拡張

漢字 1016 字種に対してデータ拡張を行い、筆記途中であるようなサンプルを作成した。例えば 6 画の漢字の場合、1 画目までのサンプル, 2 画目までのサンプル, 同様にして 5 画目まで計 5 個のサンプルを作成した。「寺」「木」「本」「土」の場合を図 2 に示す。 $m$  画の漢字に対して途中の  $n$  画までの筆記を、 $N(=10 \times n/m)$  割筆記とする。これらのサンプルは同一のサ

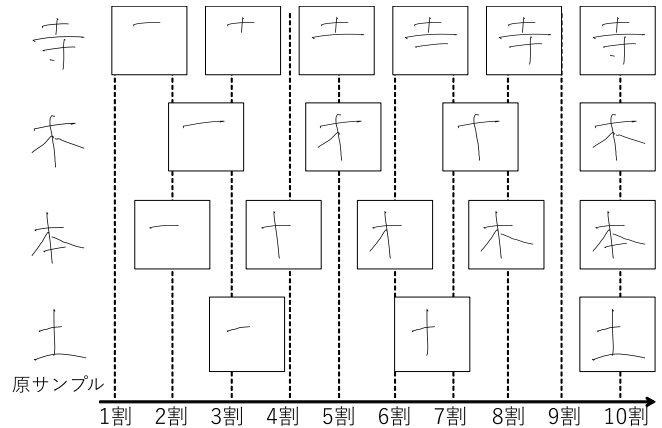


図 2: データ拡張

ンプルから生成しており、座標情報やその他の特徴量は加工しない。

## 3.2 学習サンプルの選定

漢字の 1 割筆記から 10 割筆記までの全てを学習データに含めると、特徴量が類似しているにも関わらず、異なる教師ラベルで学習される恐れがある。このような状況では、例えば「寺」の 1 画目に「寺」という教師ラベルを与えると、本来「一」と認識される文字が誤って「寺」と認識されるといった問題を引き起こす。したがって、学習サンプルとして追加する  $N$  割筆記の適切な選定が必要である。

ここで、図 3 のような木構造の文字構造ネットワークを考える。このネットワークでは、たとえば「十」「木」「二」, 「土」の 1 画目は「一」として共通のノードに対応し、その後の筆記に応じて異なる漢字のノードに遷移する。この遷移で文字が確定した瞬間の筆記率を確定筆記率とする。例えば「木」

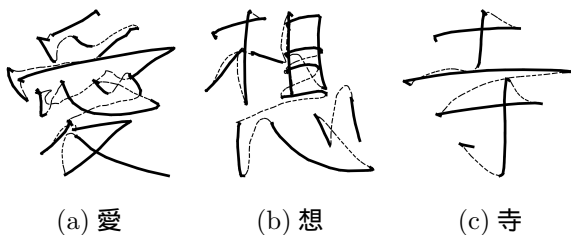


図 1: オンライン手書き文字データの例

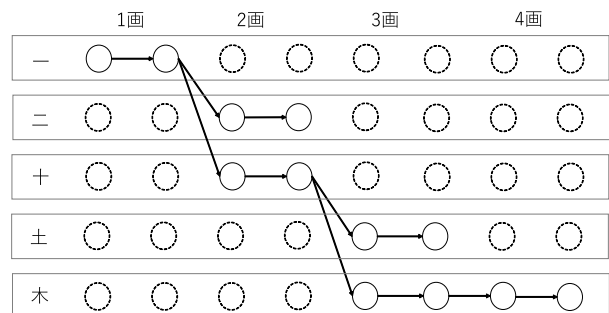


図 3: 文字構造ネットワーク

<sup>†</sup>富山県立大学, Toyama Prefectural University

の場合は 3 画目が該当する。確定筆記率別のデータの個数を図 4 に示す。確定筆記率が 50% のデータが最も多いことが分かる。よって、5 割筆記サンプルを学習データに含める。また、次に多い確定筆記率が 40% 台であるが、50% よりも低い  $N$  割筆記サンプルを学習データに含めると、誤学習が起きる。そのため、次にデータ数が多い 7 割筆記サンプルを追加する。ただし、それぞれの文字について一画ずつ書き進め、総画数の 5 割以上となる最初の筆跡を 5 割筆記サンプルとする。例えば、6 画の文字ならば 3 画、7 画の文字ならば 4 画、12 画の文字ならば 6 画まで書いたサンプルである。同様に 7 割筆記サンプルは 6 画の文字に対し 5 画、7 画の文字でも 5 画、12 画の文字に対しては 9 画まで書いたサンプルである。

### 3.3 実験方法と結果

本実験の RNN の構成は 3 層とし、ユニットには GRU を用いた。1 層目はノード数 100 個、2 層目は 300 個、3 層目は 500 個とした。学習には、50 人分の 5 割、7 割、10 割筆記 (152400 サンプル) を用いた。評価には、学習に含まれていない 10 人分の 10 割筆記 (10160 サンプル) と 5 割筆記 (10160 サンプル) を用いた。データ拡張をせず、10 割筆記のみで学習したモデルを「拡張なしモデル」、5 割、7 割、10 割筆記で学習したものを「 $N$  割筆記モデル」とする。

表 1 より、10 割筆記での評価では「拡張なしモデル」の方が「 $N$  割筆記モデル」より高い認識率となった。しかし、10 位までの累積認識率では同程度であり、最後まで筆記しても高い認識率であることが分かる。5 割筆記での評価では「拡張なしモデル」はほとんど文字を認識していない。しかし、「 $N$  割筆記モデル」は 10 位までの累積認識率で 97.1% であり、10 割筆記と同程度の認識率が得られたことが分かる。つまり、漢字を半分程度筆記した時点で、ほとんどの漢字を認識できる可能性が示された。

## 4. 全文字の筆記の総削減率

### 4.1 総削減率

学習に用いた 1016 字種全てを筆記する際の筆記量をどれだけ削減できるかについて実験を行った。画ごとに入力し、筆記

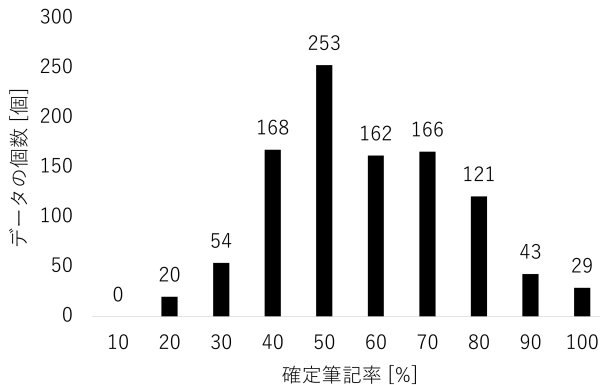


図 4: 確定筆記率別のデータの個数

表 1:  $N$  割筆記の認識率 [%] (括弧は 10 位までの累積認識率)

		評価	
		10 割筆記	5 割筆記
学習	拡張なし	97.7 (99.7)	0.09 (0.85)
	$N$ 割筆記	94.4 (99.6)	61.9 (97.1)

表 2: 各モデルにおける筆記量の総削減率 [%]

	1 位	~5 位	~10 位
拡張なしモデル	0.3	1.7	3.2
$N$ 割筆記モデル	43.3	49.3	51.4

漢字が候補に現れた時点で入力を停止する。停止時の画数を予測筆記画数とし、全体の筆記量を計算した。ここで、

$$\text{総削減率} = 1 - \frac{\text{予測筆記画数}}{\text{全ての文字の総画数}}$$

とする。漢字 1016 字種の総画数は 9571 画で、確定筆記率まで筆記した場合 5114 画となる。つまり総削減率の理論値は 46.6% である。この削減率を目標値とする。

### 4.2 実験方法と結果

10 人分の 10160 サンプルを用い、「拡張なしモデル」と「 $N$  割筆記モデル」の 2 つのモデルで実験を行った。結果を表 2 に示す。「 $N$  割筆記モデル」では、5 位までの認識候補で目標削減率を超え、49.3% の削減率を達成した。ただし、「 $N$  割筆記モデル」は漢字を筆記する際に候補として筆記したい漢字が提示されたにも関わらずその候補を見逃してしまった場合、次の 1 画を筆記することで再度候補に挙がる保証は無い。よって、実用化に向けたさらなる検討が必要である。

## 5. まとめ

本研究では、書き始めの段階で文字を認識する早期認識手法を提案した。その結果、筆記途中のサンプルを学習に加える手法は早期文字認識に有効であることがわかった。

謝辞 本研究は JSPS 科研費 24K15050 の助成を受けて行った。

### 参考文献

- [1] 栗原 一貴ほか, “音声ペン: 音声認識結果を手書き文字入力利用できる新たなペン入力インタフェース,” コンピュータソフトウェア, Vol. 23, No. 4, pp. 60-68, 2006.
- [2] 長岩 夏子ほか, “LSTM を用いた筆記途中からの筆跡補完,” 第 22 回情報科学技術フォーラム, 第 3 分冊, pp. 429-430, 2023.
- [3] 秋良 直人ほか, “ストローク HMM を用いたオンライン非目視手書き文字認識の性能評価,” 信学技法, PRMU2000-206, pp. 39-46, 2001.