研究指導 中澤 真 教授

自由手書き文字認識のための精度向上に有効な文字切り出し手法

歌川　重樹

1. はじめに

近年のインターネットの発達により，文書をデータで扱う機会が増えてきている．しかしその一方でノートやホワイトボードといった，いわゆる紙媒体に書かれる手書き文字を使用する機会も多い．これら手書きのものを電子化できればペーパーレス化につながり，紙媒体と比べて紛失や欠損の心配がなくなるだけでなく，情報の検索を効率的に管理できるようになる．

手書き文字を電子化する方法は手書きのものを手入力で電子化するほかに，手書き文章画像中の文字を機械が判読し，自動的に電子化する方法がある．これは一般的に文字認識といわれる．産業分野向けであった文字認識の技術は，文字認識ソフトやWEBサービスに適用されてオフィスや個人へと広まった[1]．文字認識のシステムは，カメラやスキャナで画像を取り込み，画像中の文字の配置などの読み込み，文字のかたまりや行の切り出した後，そのかたまりや行の中からさらに1文字ずつに切り出し，最後に切り出された文字が何であるかを認識するといった一連の流れでシステムが構成されている[2]．適切な文字認識をするには，行の中から文字を切り出すことが適切に行われることが重要である．しかし，一定の間隔で書かれ文字の崩れやつながりなどがない活字と比較して，手書きで書かれた文字の切り出しは個人の癖などの影響で精度が低くなりやすい．そのため，本研究では手書き文字の切り出し精度が向上する手法を提案する．

2. 文字認識技術について

2.1 文字認識技術の現状

現在，文字認識技術が利用されているアプリやWEBサービスは広く普及しており，スマートフォンから写真を撮るだけで簡単に電子化ができる．業務の効率化ツールとして導入している企業もあり，これによってペーパーレス化やスムーズな情報の共有に貢献している[3]．

文字認識技術の現状として，活字を対象とした場合においては非常に高い精度を達成している．また，手書き文字の場合でも帳票のように記入の枠があり，丁寧に書かれているならば精度は十分に高い．しかし，ホワイトボードやメモの走り書きといった自由に書かれた手書き文字に対しては，個人によって文字の形に差が出るため，認識結果の精度が低くなってしまうことが多い．個人が一般的に扱う機会が多いのは自由に書かれた手書き文字であるため，手書き文字に対する認識精度を高くできる方法が必要となる．

2.2 認識精度低下の原因と課題

池谷[4]は，文字の傾きや文字の間隔が文字認識精度を低下させる要因であることを明らかにした．文字が傾くことや隣接する文字間隔が狭いことで，認識の前段階である切り出しが正しく行われず，認識の精度が低下する．正しく切り出されることで高精度な文字認識の実現が可能になる．日本語の手書きでは，文字のサイズや間隔を一定に保って書くことは非常に難しく，また字形が複雑になりやすいため，これらの要因をどのように改善するかが重要である．

山本ら[5]は，日本語の手書き文字切り出しの精度向上のために機械学習を用いて，手書きによって生じる文字サイズや文字の幅，連結する場合などの文字の変動に対応した切り出し手法を提案した．この手法では，文字の切り出しの範囲を対象の文字に外接する矩形と設定し，句読点や長音などの小さく切り出されにくい文字に対して一回り大きい外接矩形と設定した学習用データを作成し機械学習をした．この学習によって，濁点のついた文字や線で囲まれた文字，画像中に文字以外の図形が混在している場合であっても文字のみを正しく切り出すことに成功し，従来よりも手書き文字一つひとつが適切に切り出せることが示された．

2.3 小さい文字に対しての切り出し精度の課題

山本らの手法により，漢字やひらがななどの一般的な文字だけでなく，長音や句読点，拗音や促音などの切り出しの精度も向上した．しかし，一般的な文字と比較すると句読点や長音，拗音や促音などの切り出し精度には改善の余地がある．サイズの小さい文字が切り出されにくい理由として，一般的な文字と比較して字形が単純であり，特徴の差が出にくいこと，書かれ方によってはほかの文字と連結していると判断されてしまうことなどが，適切な切り出しを妨げる要因になっていると考えられる．

以上のことを踏まえて，本研究では句読点や拗音などの小さい文字に対して切り出し精度を向上させる手法を提案する．

3. 切り出し実験の概要

3.1 切り出し実験の手順と評価方法

小さい文字の切り出し精度向上を図るための検証

実験の手順について図 1を使って説明する．

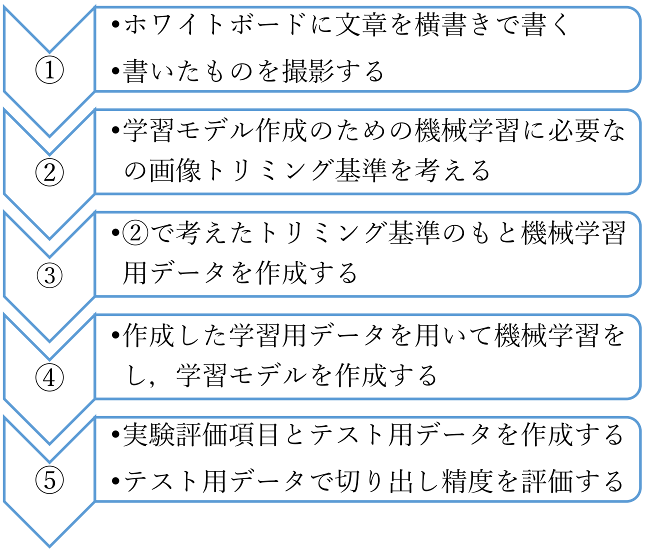

図 1 切り出し実験の手順

まず①で，実験に用いる文字画像の作成をする．画像はホワイトボードに句読点や拗音，長音を含む文章を手書きで書き，これをスマートフォンのカメラで撮影した．撮影した画像は14枚（797文字）である．

次に，切り出し精度の向上を図るためには機械学習により切り出しの学習モデルを作成する必要があるため，機械学習に用いる文字画像のトリミング基準を手順②で設定する．機械学習に用いる文字画像を学習用データと呼び，3つのトリミング基準から作成した学習用データを用いて機械学習した場合の学習モデルをそれぞれ次のように呼称する．

- 外接矩形モデル
- 最大余白モデル
- 余白2ミリモデル

外接矩形モデルは全ての文字・句読点に対して，文字に外接するぎりぎりの枠を矩形とした基準を設定して，学習用データの文字のトリミングをする．山本らは句読点や長音などを除いた文字に対して同じ条件で切り出し枠を設定している．最大余白モデルは，上下左右の文字にかからない範囲の最大の余白で基準を設定して学習用データの文字のトリミングをする．外接矩形モデルの基準と比較して，切り出しの枠にゆとりが出ることが特徴である．今回は，行間が空いているデータを用いているため，結果として上下方向の余白が大きい傾向になっている．余白2ミリモデルは，学習用データを作成するディスプレイ上の表示で横幅が17ミリ程度の文字の外接矩形から2ミリに固定して余白を基準として文字をトリミングする．他2つのモデルの基準と比較して，文字の間隔によっては図 4のように隣の文字が切り出し枠内に入ってしまうことが起こることが特徴である．余白の幅を固定することで，文字間隔が一定でない場合でも高い精度で切り出しを行えるのではないかと予想し，余白を2ミリに固定した．

手順③でトリミング基準をもとに機械学習に必要な画像の作成をする．機械学習に用いる画像は①で作成し撮影した画像を，1文字ずつ画像編集ソフトのGIMP Photo Editor[1]を用いてトリミングする．各基準で作成した学習用データを400文字分用意した．

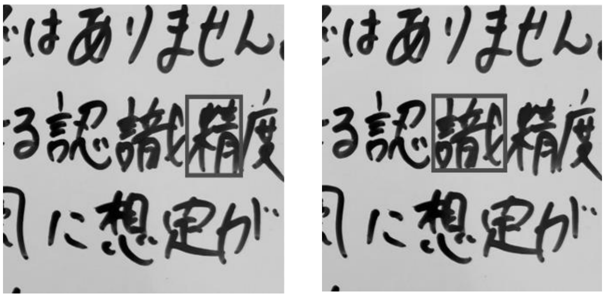

図 2 外接矩形モデルの学習用データ例

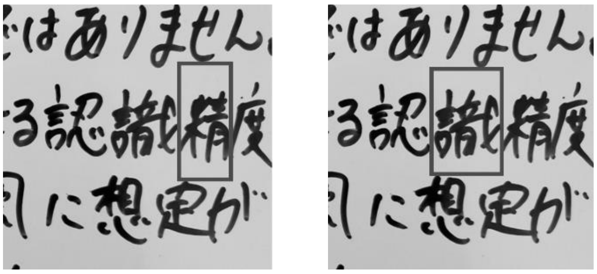

図 3 最大余白モデルの学習用データ例

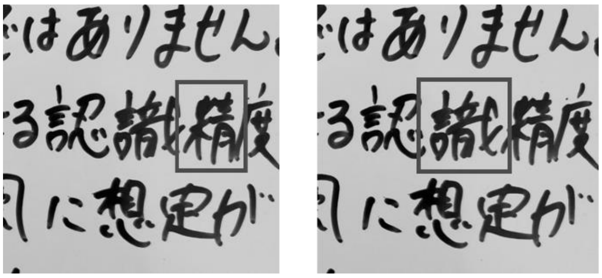

図 4 余白 2 ミリモデルの学習用データ例

④の実験に用いる切り出しをするための機械学習システムの導入について説明する．本研究では文字切り出しをするための機械学習システムとしてCRAFT（Character Region Awareness for Text Detection）[2]を使用する．CRAFTは文字とみなしたところにその文字を囲むように切り出しの枠を生成する．この枠が画面上に表示されるため，適切な切り出しができているかどうかを判断することができる．CRAFTには機械学習の機能が実装されているため，学習用の文字画像データを用いて切り出しの精度をより高めることも可能である．そこでCRAFTに③で作成した学習用画像を入力して機械学習をさせた．

最後に⑤の実験の評価について説明する．機械学習後のモデルの切り出しテストとして，学習用デー

[1] https://forest.watch.impress.co.jp/library/software/gimp/

[2] https://github.com/clovaai/CRAFT-pytorch

タとは異なる手書き文字文章の画像を20枚（1438文字）用意してテスト画像とする. 本実験では, 先行研究の再現をしたモデル, 外接矩形モデル, 最大余白モデル, 余白2ミリモデルの切り出し精度の結果を比較する. テスト画像20枚中に含まれる文字をそれぞれ句読点,「ゃ, ゅ, ょ」などの拗音と「っ」の促音, カギ括弧などの記号, その他の一般の文字に分類した. 切り出しの成否については, 生成された文字枠1つに対して, その枠内に1文字が正しく切り出されている場合を切り出し成功とし, 隣の文字と連結して枠が生成されてしまう, 画像中に実際に存在している文字があるにもかかわらず枠が生成されない場合などは, その文字の切り出しはいずれも失敗とする. 結果はそれぞれの文字種別ごとに, 成功数をその文字種別の出現数で除したものを成功率として算出した. この結果を表 1に示す.

表 1 モデル別の切り出し精度の比較

内訳	総数	先行研究再現	外接矩形	最大余白	余白2ミリ
句読点	74	69%	62%	86%	66%
拗音促音	51	71%	76%	84%	63%
「」などの記号	54	70%	74%	91%	72%
その他	1259	92%	88%	65%	63%
全体	1438	89%	86%	68%	64%

3.2 切り出しの結果と評価

山本らの手法を再現したモデルでは, 表 1で示した全モデルの中で総合的には最も高い精度であった. しかし, 既に述べたように句読点や拗音促音などの文字の切り出し精度にはやはり課題があることも確認できた. 外接矩形モデルでは, 先行研究に似た結果となり, 問題点の改善は見られなかった. 最大余白モデルでは, 句読点や拗音促音, 記号に対してほかのモデルよりも切り出しの精度が向上したが, その他一般的な文字に対しての精度が大きく低下してしまったため, 全体の精度が大きく下がるという結果となった. 余白2ミリモデルは, 各項目の切り出しの精度が低くなり, 手書き文字に対応できていないことが確認された. CRAFTが実際に切り出しを行った画像として, 外接矩形モデルの結果を図 5, 最大余白モデルの結果を図 6に示す.

図 5に示した外接矩形モデルでは, 拗音や漢字などに対して正しく切り出しができており, 総合的に精度は高い. 一方で, 句読点や拗音などに対してはほとんど枠が生成されず, 切り出しができていない. また, 1行目の「に」のように文字の中で生じている空白部分を文字と文字の境目であるとCRAFTが判断し, 誤った切り出しも発生している.

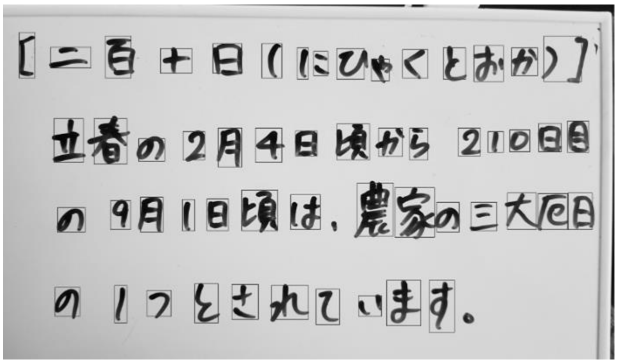

図 5 外接矩形モデルによる切り出し

図 6に示した最大余白モデルでは, 切り出しの枠が文字に対して大きく生成されている. この方法は句読点に対して切り出しが反応でき, 外接矩形モデルでは分離して切り出されてしまった文字も適切に切り出しができている. しかし, 隣の文字に干渉してしまったり, 連結してしまったりする箇所が見受けられ, 全体的に切り出し精度は低下している.

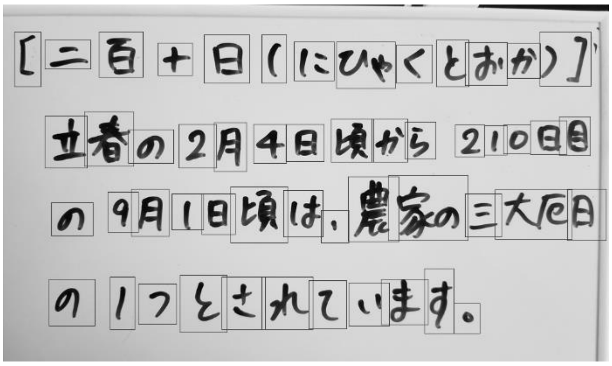

図 6 最大余白モデルによる切り出し

3.3 精度が向上しなかった考察と新たな実験

前節の3つの手法で改善には至らなかった原因として, 1モデルのみの機械学習では手書き文字特有の形の崩れや文字の連結などの様々な特徴に対応しきれなかったからではないかと推察される. しかし, 外接矩形モデルは全体的には高い切り出し精度を維持し, 最大余白モデルでは句読点などの切り出し精度が向上した. 以上のことを踏まえて, この2つのモデルの学習用データを掛け合わせることによって, 総合的な切り出し精度の向上が期待できるのではないかと考えた. この2つのモデルを掛け合わせてできた新たなモデルをハイブリッドモデルとする.

これまでの実験を踏まえ, 外接矩形モデルの学習用データを200枚, 最大余白モデルの学習用データ200枚の計400枚の学習用データを用いてCRAFTに機械学習させた. 学習用データを200枚ずつにした理由は, 合計の学習用データの量をそろえることで学習データ量の差が結果に影響を及ぼさないようにするためである. 学習用データ, テスト画像は前節までの実験と同じ画像を使用し, 評価方法もこれまでの実験と同じ項目・内容で評価する.

4.　ハイブリッドモデルの結果と評価

4.1 ハイブリッドモデルの結果

ハイブリッドモデルによる学習の結果を表 2に示す．句読点や拗音，促音など先行研究では精度が低くなっている項目の精度向上に成功していることが確認できた．また，その他の項目でも先行研究と同程度の切り出し精度となり，これまでのモデルでは達成できなかった総合的な切り出し精度の向上が達成できた．

表 2　ハイブリッドモデルの切り出しの精度

文字種別	出現文字数	先行研究再現（200文字学習）	先行研究再現（400文字学習）	ハイブリッドモデル
句読点	74	69%	70%	82%
拗音促音	51	71%	71%	86%
「」などの記号	54	70%	71%	87%
その他	1259	92%	92%	89%
全文字	1438	89%	89%	88%

ハイブリッドモデルによる切り出しを行った実際の画像が図 7である．この結果から，かなり正確に文字に対して切り出しの枠を生成していることが確認できる．文字と文字との間隔が狭い箇所でも隣に干渉することなく切り出しができている．句読点に対しても切り出し枠が正しく生成されており，記号の切り出しも問題がないことが確認できた．

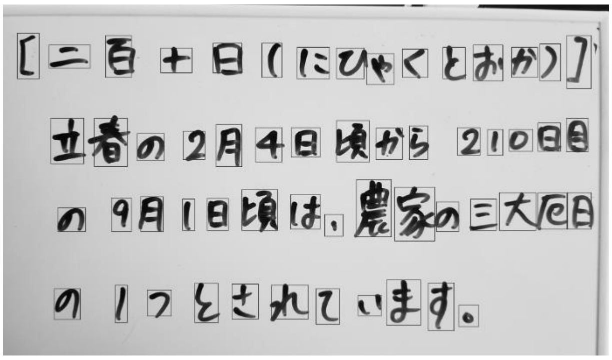

図 7　ハイブリッドモデルによる切り出し

4.2 精度向上理由の考察

外接矩形モデルと最大余白モデルの両方の学習を掛け合わせたハイブリッドモデルの精度が高くなった理由として，1つのモデルのみの機械学習では対応しきれなかった1文字1文字の正確な切り出しと句読点や記号，拗音促音の切り出しの両立が，2つのモデルの機械学習をしたことで可能になったからではないかと推察される．図 7で示した切り出し結果では，その他に分類される文字には矩形が外接矩形のように生成され，句読点などに対しては外接矩形より一回り大きく枠が生成されている．その他の一般的な文字に対して外接矩形に近い形で切り出しが行われたことは外接矩形モデルの機械学習によるものであると考えられる．また，句読点や拗音・促音などに一回り大きく生成されている切り出し枠は最大余白モデルの機械学習によるものであると考えられ，2つのモデルの機械学習を掛け合わせたことでそれぞれの文字種類に対応し，総合的に精度が向上したといえる．

5.　まとめ

本研究では，文字を外接矩形でトリミングした学習用データと，上下左右の文字にかからない範囲の最大の余白で文字をトリミングした学習用データの2種類を用いて機械学習させることで，手書き文字に生じる文字の連結やサイズの違いに対応しつつ，特徴が出にくい句読点などの小さい文字に対しても高い精度の切り出しができるようになることを示した．日本語は漢字やひらがな・カタカナなど文字の種類が他言語と比較して非常に多く，形や大きさなど切り出し精度向上に対応しなければならない要素が多い．今回の手法では，一般的な文字だけでなく句読点や拗音・促音，さらにはカギ括弧などの記号にも対応し，切り出し精度の向上が達成できた．

問題点として，カギ括弧などの記号や句読点が文字に近すぎる場合は，近くの文字と連結して1つの文字として切り出してしまう場合が複数見受けられた点が挙げられる．近すぎる文字や句読点の切り出し精度を向上させるためには，新たなトリミングの基準を設定して学習用データを作成する必要があると思われる．今後は，切り出しが不正確であった記号や句読点に対してのさらなる改良に取り組む予定である．

参考文献

[1]　黒沢由明,入江文平,水谷博之,登内洋二郎,“実社会での利用が広がる文字認識技術”，情報処理, vol.51, No.12, pp.1530-1537, 2010.

[2]　坂井邦夫，入江文平，水谷博之，“知能ロボットの技術：人工知能からのアプローチ（前編）：3，文字・文書の認識と理解”情報処理, vol. 44, No. 11, pp.1123-1129, 2003.

[3]　文字認識の技術水準について，メディアドライブ,https://mediadrive.jp/technology/techocr07, (参照 2022-6-29).

[4]　池谷友宏，“自由手書き文字の認識精度低下の要因分析”，会津大学短期大学部産業情報学科経営情報コース, 2018 年度卒業研究論文要旨集, pp.53-56, 2018.

[5]　山本祐美,本郷仁志,森吉弘,城和貴，“ホワイトボードからの文字抽出手法の検討”，情報処理学会研究報, vol.2019-MPS-126, No.14, pp.1-4, 2019.

[6]　志久修,手島裕詞,内田誠一，“傾斜文字認識のための正規化方法”，電子情報通信学会論文, vol.J-100D No，10, pp.902-906, 2017.

[7]　日本語 OCR はなぜ難しい？NAVER のエンジニアが語る、テキスト検出における課題解決策, logmiTech, https://logmi.jp/tech/articles/322529,（参照　2023-1-19）.

[8]　JetsonNano と CRAFT で日本語文字検出, Qiita, https://qiita.com/KTake/items/dbce1e7361fe0f03139a,（参照　2023-1-20）.