

[フェロー記念講演] あなたがいま読んでいるものは文字です

— 画像情報学から見た文字研究のこれから

内田 誠一[†]

[†]九州大学大学院システム情報科学研究院 〒819-0395 福岡市西区元岡 744

E-mail: tuchida@ait.kyushu-u.ac.jp

あらまし 本フェロー講演では、画像情報学から見た文字研究のこれからについて論じたい。本講演を通して、第一に文字は画像情報学的に極めて面白い研究対象であり、まだ着手されていない研究課題が多いことを伝えたい。それを通して第二に文字研究は文字認識だけではないことを伝えたい。新たに文字研究を目指してくださる方が増えれば本望である。

キーワード 文字にしかできない研究, 文字だからやりやすい研究

What you read are thirty characters

— Open problems in character image analysis research

Seiichi UCHIDA[†]

[†] Faculty of ISEE, Kyushu University 744 Motoka, Nishi, Fukuoka, 819-0395 Japan

E-mail: tuchida@ait.kyushu-u.ac.jp

1. ま え が き

本講演では、「画像情報学における文字研究のこれから」について述べたい。フェロー記念講演の一般的な目的は、受賞者が自分の過去の研究を振り返りながら、達成した学術的貢献や、それに至るプロセスを紹介することで、後進の研究者に役立ててもらふことにある（と勝手に想像している）。しかし、筆者の場合、何か大きな業績を為したというよりは、特に文字に関する研究を「しつこく続けてきた」ことが評価されたように感じている。もし本当にそうであれば、このフェロー記念講演の場において、文字研究の面白さやこれからの課題を伝えることが、筆者にしかできない恩返しになるはずだと考えた。ご理解いただければ幸いである。

本講演の Take-home messages は次の二つである。すなわち第一に「文字は画像情報学的に極めて面白い対象である」こと、そして第二に「文字研究 ≠ 文字認識研究」^(注1)である。第一のメッセージについては、次節以降、幾つかの視点に分けてそれぞれ紹介する。実際、深層学習により文字検出・認識の性能が劇的に向上したいまだからこそ新たに組みめる文字研究群も

多い。これらの紹介を通して、多くの人々が持っている「文字研究=文字認識研究」という誤解を解きたいというのが、第二のメッセージである。

もちろん歴史ある文字認識研究は、今でもそして今後も極めて重要である。例えば、古文書認識は情報学だけでなく歴史学的にも大きな意義がある。RPA(Robotic Process Automation)、そして自動採点等でも注目されている自由手書き認識などは、これからの産業上極めて重要である。帳票認識や論理レイアウト認識も、古典的な課題でありながら、本質的な困難性は依然解消されていない。OCR と自然言語処理との連携も推進すべきであろう^(注2)。これらユニバーサルな OCR 実現に向けた文字認識研究については、今後も引き続き高精度化を目指すべきであると考えている。

しかし、それら広く知られた認識課題以外にも、文字研究には様々な取り組むべき興味深い課題があるのも事実である。(厚顔無恥な) 筆者は、このことを“Beyond 100%” [3] と称して、従前より国内外で機会あるごとに述べてきた。今回多少の私見

(注1) : もちろん「MNIST の認識率を 100% にする研究」ではない。

(注2) : Paper Gestalt [1] の進化系として、「数式・論理レイアウト・図表を認識し、さらに自然言語処理を経て、論文の合否判定を実現してくれるような OCR システムができれば...」と思っている研究者も多いに違いない。

が許される（と勝手に想像している）このフェロー記念講演という場においては、この話題をやや詳細に述べさせて頂きたい。なお、以下で述べるアイディアはすべて筆者によるものではなく、文献 [2]^(注3) や、筆者の協働研究者との議論によるものも多いことを、お断りしておく。

2. 画像パターンとしての文字

画像情報学で一般的に使われる顔画像などに比べて、文字は画像パターンとして様々な特異な性質を持っている。本節では、これらの性質について考察する。

2.1 Plain な画像である

文字は Plain な画像である。すなわち、多くは均質な背景上に描かれた単一色のパターンであり、従って二値画像である。そして複数の線分（ストローク）で構成されており、従って連続的なエッジを持つ。さらにこの線分幅はほぼ一定である。他にも、高々 32×32 画素程度でも表現しうる点や、二値画像として見た場合の 1 文字当たりの連結成分数は高々 10 程度である点、奥行きのない平面上のパターンである点が挙げられる。このような特殊なパターンであるから、その処理に当たってはその性質を強く意識すべきと考え、単純な文字画像処理（変換）に限っても、依然様々な研究課題があるように考えられる。

2.2 あらかじめ決まったクラスが存在する

あらかじめ決まったクラスが存在することも、文字の特異性の一つである。例えば英語大文字であれば、26 クラスと決まっている。3. で述べる通り、これは文字がコミュニケーションのための離散通信符号であることに因る。文字は、それがコミュニケーションに使われる限り、必ず事前に定義された有限離散個の記号集合の中から選ばれる。すなわちクラスが明確に規定されている。

この事実は、文字がパターン認識の対象として極めて有利な性質を持つことを意味する。実際、一般物体、例えば机や椅子の識別の場合、それら 2 クラスは本質的に識別できない可能性がある。なぜなら機能を持った椅子の存在を否定できないためである。別の言い方をすれば、「家具認識」をする場合、どのようなクラスをいくつ準備すべきかは、自明ではない。

2.3 配色は限定的である

前述のように文字は 2 値画像であるが、これは白黒に限定されるわけではない。様々な色が利用される。ただし、文字色と背景色の組み合わせは完全にランダムではない。情景内の文字について以前調査したところ [4]、(1) 明暗のコントラストがある、(2) 彩度のコントラストがある、(3) 有彩色の組み合わせは主として「赤と黄」に限定される、ことがわかった。特に (2) は重要であり、文字色が背景色のどちらかは無彩色であることが多いことを意味する。従ってどちらも有彩色というケースは少数であり、あったとしても (3) で述べた通り配色は極めて限定的である。こうした事実からいわゆる視覚的顕著性も高くなることがわかっている [5]。要するに、文字は読んでもらわな

れば意味がないので、目立つように配置されているということであろう。

2.4 人為的にデザインされたパターンである

全ての文字記号（‘A’ や ‘愛’ など）は、人間が自身のコミュニケーションのために、自らデザインした視覚的な通信符号である。その際、通信誤りすなわち読み間違いが起きにくいように、慎重にデザインされたと予想される。このことを岐阜大学名誉教授の山本和彦先生は「文字は脳によって淘汰されたパターンである」と表現している [6]。

読み間違いが起きにくいパターン集合をデザインするというのは、そう単純な話ではない。様々な変形やノイズが混入しても読み誤らないためには、巧妙な工夫が必要である。例えばある民族が、線分を $0^\circ, 1^\circ, 2^\circ, \dots$ と傾けたものを、それぞれ $0, 1, 2, \dots$ を表現する文字（数字）記号としてデザインしたとする。この場合、見る角度によって数字が変わってしまい、非常に不便である。筆者の知る限り、こうした数字記号体系を持った言語は無く（例えば文献 [7, Sec. 69]）、人類の英知を感じる。逆に考えると、複雑に見える漢字や西夏文字^(注4)でも認識しやすくデザインされていると考えられる。

2.5 繰り返し構造を持つパターンである

文字は単独で使われることは少なく、多くの場合、文字列として存在する。文字列の中には 1 文字を単位とする繰り返し構造が存在し、文字幅がほぼ一定ならばその繰り返し周期も一定となる。さらに複数行になった場合、行単位での繰り返し構造も存在することになる。言わば二次元の周期性を持ったテクスチャ状のパターンを構成することになる。一方、一般的な CNN による画像認識は局所構造の aggregation が基本である。周期性のような大局的な構造を明示的に意識した構造に改変することは可能だろうか？

3. 通信符号としての文字

3.1 誤り訂正符号としての文字

文字はコミュニケーションのための離散通信符号であり、そして極めて優秀な誤り訂正符号でもある [2]。下手な手書き文字でもある程度読める。すなわち多少の変形があっても、復号可能である。もちろんパラメトリックな幾何変換にも頑健であり、射影変換程度であればほぼ問題なく読める^(注5)。2.4 で述べた「傾いた線分による数字体系」を用いなかったのも、この幾何変換に対する誤り耐性のためと考えられる。さらに、ドットノイズや線状ノイズを加法的に与えても読める [2]。筆者は詳しくないが、ノイズの中にあっても文字としてのまとまりを認識するという、人間の認知機能が背後にあるのかもしれない。

3.2 文字と非文字の境界

文字と非文字の境界把握も、極めて興味深い問題である。ど

(注3)：「パターン認識の今後の課題」が列挙された書籍。1994 年発行でありながら、2019 年現在でも非常に参考になる議論がなされている。

(注4)：西夏文字は、一見漢字に似ているが、象形文字ではないので漢字より解読は複雑である。すでに消滅した記号体系であるが、unicode 化されており、次の URL で一覧できる。<https://www.unicode.org/charts/PDF/Unicode-9.0/U90-17000.pdf>

(注5)：ただし、裏返った「ざるそば」の職が「ぢるそば」に読めてしまうことはある。

ここまで変形が進めば「文字ではない」すなわち復号する必要がないとしてよいのだろうか？ 情景内の文字を検出・認識する課題が扱われるようになって、この問題が顕在化している。筆者の研究室で ICDAR2013/2015 の情景内文字データセットを構築した際、どこまでを文字として Ground-truth を与えるかは、常に問題になった。特に悩ましいのは小さく写った値札や本の背表紙で、「これは間違いなく文字なのだが、小さすぎて読めない」箇所である。同データセットではこうした曖昧な箇所を don't care とする妥協策を採ったが、より本質的な解はないだろうか？ リジェクトの問題を文字独特の立場から吟味することはできないだろうか？

なお、筆者は非文字に二種類のタイプがあると考えている。すなわち「文字の形に似た木の枝」のような「紛らわしい型」と、「(‘木’の字を上下反転させたような) どうみても文字の一種なのだが、世界中のどこでも使われていないパターン」という「未利用型」である。特に後者の存在は「文字だからこそ」起こる検討事項に思える。

4. 時系列パターンとしての文字

4.1 文字のマルチモーダル性

文字の生成法は元来手書きが主流であり、そして手書き文字はペン先座標の時系列として表現できる。すなわち手書き文字は、「画像」として観測されつつも、同時に「時系列」性を持つという、本質的にマルチモーダルなパターンである^(注6)。漢字のような多画文字では状況はさらに複雑である。すなわち各ストローク内の筆順だけでなく、ストローク集合も何かしらの順序（いわゆる書き順）に従うので、ある意味で二重の時系列性を持つことになる。さらに画と画の間にペンアップ動作という不連続性が入る。

4.2 文字筆記の非マルコフ性

時系列パターンとしてみた文字の特殊性の最たるものとして、個人的には非マルコフ性を挙げたい。ジェスチャ認識などの時系列パターン解析においては、HMM や RNN を利用するためにマルコフ性が仮定されることが多い。ところが、文字の筆記プロセスはマルコフ性だけでは表現できない。実際、数字の‘0’を書く時、マルコフ性の仮定の下では、始点終点という時間的に離れた2点間の関係を制御できないので、閉じた0を書くことは保証できない。このように手書き文字生成の過程を考える場合、マルコフ性で表現しうる局所的構造と、全体のバランスといった大局的構造の両方をモデル化する必要がある。

4.3 文字筆記のマルチスケール性・多目的性

さらに、手書きゆえに、人間の腕や手の関節構造や運動能力といった制約条件から逸脱することは不可能である。従って、Reaching モデル（例えば [8] や [9]）や Kinematic theory（例えば、Sigma-lognormal モデル [10]）を考える必要がある。以上より、手書き文字の総合的な筆記モデルを構築するためには、

局所的構造のためのマルコフ性、メゾ構造のための骨格・運動モデル、そして大局的構造のための視覚フィードバックが必要になると考えられる。

以上のことから、文字のデザイン過程が、Generative Adversarial Networks のような仕組みで表現できるようにも感じる。すなわち、筆記モデルに従う文字の Generator と、文字として読めるための Discriminator が戦った結果、いまのような文字体系ができた可能性はないだろうか？ またその際の Discriminator の loss 関数にはどんな要素が必要であろうか？ 少なくとも、様々な変形・ノイズにする耐性および他の文字との弁別性能の二つは必要ならば、多目的最適化となる。他には何が？

4.4 文字筆記の発達過程

文字筆記については、幼児および児童の描く線画（スクリブル, scribble）の発達過程に関する様々な考察も参考になるかもしれない。文献 [11] によれば、2 歳児による線画は、視覚的フィードバックを用いずに筋肉運動のみで作られ、それらは 20 種類の基本スクリブルに分類されるという。そして 3 歳ぐらいになり、視覚フィードバックが適切にできるようになると、基本スクリブルを組み合わせた「ダイアグラム」を描画できるようになる。その代表例として矩形や円形、三角形、そして十字が挙げられている。そしてダイアグラムを組み合わせたコンバインやさらに複雑なアグレゲイトを構成できるようになるという。こうした考察は、上記の局所、メゾ、マクロの話と関係するようにも思える。

4.5 文字筆記からの筆記者情報推定

時系列としての連筆情報は、筆記者の特性や状態を端的に表す点でも重要である。サインによる認証は、筆記の個人性を利用したバイオメトリクス応用である。他にも筆記に基づく男女推定、年齢推定、感情推定、性格推定、そしてパーキンソン病や失読症の判定が試みられている。筆記内容だけでは見えてこないことを、筆記形状やその他連筆情報が補完してくれる可能性がある。

5. ヴィジュアルデザインとしての文字

5.1 多様なフォントの存在

同じ‘A’という文字記号だけを見ても、膨大なフォントのバリエーションが存在する。通信記号として誤りなく情報を伝えるのであれば、常に同じ‘A’の形状を使うのが最も高効率なはずで、フォントとして形状にバリエーションを持たせる必然性はない。それにも関わらず、人類はこれまでに膨大な数のフォントをデザインしてきた。

これには幾つかの理由がありうると考えられる。第一には、より高い可読性の希求である。すなわち、石板、紙、ディスプレイなどと提示メディアが変化する中、より読みやすくするために字形が整えられてきた可能性がある。昨今の UI フォントもその流れである。第二には、機能性の付与である。磁気読み取り用の MICR フォントや、機械読み取りを阻害するための Captcha、失読症向けのフォントが、その例である。第三には、人間のデザイン欲求がある。他の芸術同様、デザイナー自身

(注6)：さらに文字は音として声に出して読めるので、音のモダリティも持つ。さらに文字（列）が表現する意味のモダリティもある。これらのことから、文字は「形・音・義」を持つと言われる。

の美学に基づいて、新奇な表現を目指した結果としてのフォントも多々あると思われる。第四の、そして筆者が最も注目するのは、次節 5.2 で述べるように、特定印象を与えたいという理由である。

5.2 フォントと印象

フォントを適切に利用すれば、文字で書かれている内容に加え、特定の印象を読む人に与えることが可能である。例えば、高級ブランドロゴに適したフォント、菓子のパッケージに適したフォント、銀行のロゴに適したフォント、警告に適したフォント、ホラー映画の広告に適したフォント、等々と言われると、読者の脳裏にはそれらしいフォントが浮かび上がるのではないだろうか？（おそらくそれはビジネス文書に使われるフォントとは異なるであろう。）また、フォントを変えることで、同じ「笑う」という単語のまま、微笑、爆笑、冷笑、嘲笑等を表現しうること [12, p.97] も、想像できるであろう。

フォントデザイナーとタイポグラファーは、このような「フォントが与える特定印象」のエキスパートである。すなわち、読者に非言語的な情報としての特定印象を与えるべく、フォントデザイナーは多様なフォントをデザインし、またタイポグラファーは適切なフォントを選んで利用している可能性がある。フォントデザインと印象の関係については、主観評価に基づいた研究が 1920 年代から行われており [13], [14], またユースケースを紹介した書籍も多くある [15], [16]。しかし、その関係を大規模データに基づいて客観的に定量化した研究例は、筆者の知る限り、ほとんどないようである [17]。なお、フォントと印象については、フォントの形状だけでなく、色にも注目すべきと考えている。実際、文献 [17] で実験的に検証したように、タイポグラファーは明らかに状況に応じて色も使い分けている。

5.3 フォントデザイン研究のさらなる展開へ

フォントデザインと印象の関係がわかったとすると、新たな研究が展開できるだろう。すなわち、なぜそのようなデザイン要素がそのような印象を与えるかを解明する必要が生ずる。これには先天的要因と後天的要因があるように筆者は想像している。前者は例えば「太い＝安心感」といった物理的感覚によるものや、bouba-kiki 効果のような形状が音やリズムを連想させるという共感覚（シナスタジア） [18, p.87] によるものもあるだろう。後者は、例えば明朝体で印字された新聞を読み続けたので「明朝体＝信頼感」という印象を持った、といった状況である。当然ながら、これらの検討には、認知心理学研究との協働が必要になるはずである。

フォントデザインと印象の関係解析をする際に、文字で構成される単語や文が意味を持つことを利用すべきと考える。すなわち、印象という主観的なものの代わりに、単語が表現する意味を利用することを考える。例えば、「甘さ」に関して類した意味^(注7)を持つ「チョコレート」「キャンディ」という単語が、どのようなフォントで印字されているかをみることで、甘さを表現するために相応しいフォント形状の傾向がわかるだろう。さ

らに、文字の配色と意味との相関関係を調べることも価値があると思われる。

フォントの形状と印象の相関解析には、漫画における描き文字によるオノマトペ表現（「ドーン」「ガシャーン」など）の形状解析が役に立つかもしれない。オノマトペは音や質感をダイレクトに表現しているため、より直接的な相関を見つけやすい可能性がある。ただし、描き文字は検出や認識も難しいので、挑戦性の高い研究になるだろう。また、「キャラの声をフォントで再現する方法」という雑誌記事もあり [19]、フォントデザインを巡る様々なモダリティの相関には依然多くの検討課題があるように思える。企業のブランドロゴにも文字が含むものが多いので、デザインされたロゴの解析や、ロゴの生成過程（[20] は参考になる）の解析も、デザインと印象を広くとらえるための研究課題として好適であろう。なお、こうした飾りの多いフォント（いわゆるデザイン書体）ではなく、ビジネス文書における代表的な Arial(サンセリフ体)と Times New Roman(セリフ体)についても、どちらが説得力があるかといった主観の評価も行われている [21]。

6. ラベルとしての文字

筆者は、我々の身の回りにある文字情報がその役割の観点から「メッセージ」と「ラベル」の二種類に分かれると考えている。メッセージとは、文字情報そのものが独立で意味を持つもので、スマートフォン上に提示されているニュースの文字や、書籍紙面上の文字が相当する。一方ラベルとは、その文字列が添付されている物体もしくは周囲を明確化するためのものであり、従ってその場を離れてはその価値を失うものである。ワインの瓶に貼付されているラベルはまさにその例であり、その瓶の中に何が入っているかを明確化している。他にも商品パッケージ上の商品名、ビルの看板や道路の案内標識、ドアについた「押/引」、エレベータの階数ボタンなどはラベルとするのが自然であろう。

文字がラベルとして機能することは、文字の重要な性質を表現している。すなわち、文字にはセマンティックギャップがないという点である。例えば「ワイン」という文字は、特定の飲料をダイレクトに指定しており、曖昧性がない。一方、ワインの瓶は、ラベルが無ければそれが何であるのか理解するのは困難である。要するに、こうしたセマンティックギャップを埋めるために、文字が貼付され利用されているのである。

この文字のラベル機能については、物体認識への援用などで徐々に注目されてきているようだが、まだ様々な興味深い研究課題が残っているように思える。まずメッセージとラベルの識別がある。完全に区別できない場合もあるだろうから、問題設定には工夫が必要であろう。他にも、どのような事物がラベルにより明確化されているのか、画像内のあるラベルが明確化している領域はどこか、ラベルを利用することで image captioning の情報量がどの程度増えるのか [22]、などである。情景内文字を用いて質問に答える Visual Q&A [23] というコンペティションも開催され、単なる情景内文字認識の一步先に進んだ課題として歓迎したい。

(注7)：ここでいう「意味」とは、厳密なものではない。「単語埋め込みの結果似たベクトルになる 2 単語は、似た意味を持つ」程度を指す。

7. メッセージとしての文字

メッセージとしての文字情報も、興味深い研究課題を与えるに違いない。我々は文字を経由してどのような情報を日々得ているのだろうか？大阪府立大学教授の黄瀬浩一先生がよく引用されている“*You are what you read*”というフレーズがあるように、我々の知識の多くは文字を経由して得られたものである。従って、情景内文字認識技術を活用して、日々読んでいる文字のすべてを自動記録しておけば(Reading-life Logと呼ばれる)、それを解析することで、我々の知的活動の定量化ができるだろう。どのようなメディア、時間、場所から、どのような知識が得られるか？またそのトレンド、頻度、情報量、個人差、個人内変動、成績との関係はどうか？さらに極めて重要な「読後の行動への影響」など、様々な研究課題が広がっている。もちろん、ログ自体も備忘録などに利活用可能である。

8. メディア変換のハブとしての文字

最後に、画像情報学というよりは自然言語処理に近い分野の話になるのかもしれないが、メディア変換のハブとしての文字の重要性について触れたい。まず、様々なメディアに対するパターン認識の結果は、一般にクラス「名」、すなわち文字列で表現される。すなわち、音声認識は音(波形)を文字に変換する処理であり、画像認識は画像(2次元信号)を「猫」のような文字列に変換する処理である。画像キャプションは、画像を文に変換する処理である。そして、文字からメディアに変換する処理もある。すなわち、音声合成は文字列を音声信号に変換する処理である。また昨今では説明文を与えればそれに合致した画像を合成する技術(例えば[24])も検討されている。

要するに、メディアの変換の際には、文字を経由することが多い。すなわち、多くのメディアは文字に変換でき、逆に文字は多くのメディアに変換しうる。そこで、異なるメディアであっても、同じ文字列として表現しうることを理由に、関連付けて扱うこともできる。このことでゼロショット学習の拡張系のような様々な研究が生まれる可能性がある。さらにこうした研究に、画像情報学に見た文字の特異性を活用することも可能かもしれない。

9. あとがき

9.1 「対象の性質の見極め」について

筆者が文字認識研究を始めた1990年代後半、様々な文献や諸先輩のご指導から叩き込まれたことは、「対象の性質を見極めること」であった。実際、その少し前、文字認識研究の全盛期のころには、高認識率を与える様々な特徴が、文字という対象の性質を見極めた研究者によって提案されていた[25]。それらには認識性能の優劣はあったものの、いずれも何かしらの視点で文字の特徴をつかんでいたように思われる。

一方、いまや表現学習を手に入れた我々は、こうした見極めの必要がなくなったように見える。深層ニューラルネットワークを用いれば、文字と一般画像の区別なく、適切な特徴が学習され、高い認識性能を手に入れることができる。文字研究に対

する参入障壁が下がったということで、筆者はこの状況を大いに歓迎したい。(文字認識系企業が急増しており、産業界は大変そうであるが...)

ただし今後、次の三つの理由により、特徴抽出を含めた「対象の性質の見極め」が再び重要になってくると考える。第一は、基本的な深層ニューラルネットワークの性能をさらに向上させるには、結局ドメイン知識が必要になるのではないかと、いう予想である。特にデータが大量に得られない特殊文書の場合など、「大規模データ × 深層学習」の定石に載らない場合には、ドメイン知識に基づく prior が重要になるだろう。第二は、explainability である。深層ニューラルネットワークの挙動を解釈するためには、ネットワーク内部で対象のどのような特徴を捉えているのかを理解し、挙動の妥当性を納得する必要が生じるだろう。

第三に、深層学習で獲得した特徴が、我々が暗に想像しているものはない可能性がある。筆者のグループでは活字数字と手書き数字(数字なので全10クラス)を混ぜてCNNを学習し、同じ数字(例えば‘0’)活字と手書きがCNNのどのレイヤで統合されるのかについて検証したことがある[26]。その結果、特徴抽出層では全く統合されず、すなわち同じ‘0’であっても活字と手書きでは異なるクラスのように扱われており、最後の全結合層で「これらは同じく‘0’」と識別されていた。識別精度は非常に高く、工学的には全く問題ない。しかし、我々人間は活字と手書きで異なる特徴抽出するように学習するのだろうか？もちろんその可能性もあるが、もし我々が未知の文字について手書きと活字を対応付けられるならば、特徴抽出の段階でもある程度の融合は行われているようにも思える。識別が非常に困難そうな物体A,Bの画像が見事に100%分離できて喜んでいたら、実は物体AとBで背景の明るさが違った、というのはよくある話である(深層学習に限った話ではない)。

筆者は、文献[27]を時々思い出して自戒している。1982年発表の同文献では、認識に有効な文字の位相的特徴を設計するのに先立ち、手書きの漢字「田」のループ数(閉じた白領域の数)の分布を調べている。活字のように書かれた「田」であればループ数は最大値4となる。内側の「+」と外側の「口」の接触箇所が減るとループ数も減少し、さらに外側の「口」が閉じていない場合には最小値0となる。調査の結果、驚くべきことに0から4までほぼ同数であることが判明した。そしてその事実により、ループ数の利用では分類が困難であると述べている。対象の性質を見極めるには、結局はこのような地道な観察が重要である。パターン認識研究者として、今だからこそのような態度を忘れてはならないと思っている。

国の施策としてデータサイエンスが喧伝されているように、情報系をハブとした様々な学術分野の連携・協働が今後加速することは間違いない。その際、実データ解析のエキスパートである我々パターン認識系の研究者は、情報系の中で最も重要な役割を担うように感じている。そのためにも、実データに潜む傾向や特徴を見抜く力を我々は失ってはならない。Factfulness[28]やEBPM(Evidence-Based Policy Making)を推進するのは我々なのである。このような対象の性質を見抜く

ための基礎力を培うためには、文字研究が好適であると、筆者は確信している。さらに、対象の性質だけでなく、時にブラックボックスと揶揄される機械学習の性質を見極める上でも、やはり文字ほど適した題材も無いのではないだろうか。

9.2 文字にしかできない研究, 文字だからやりやすい研究

最後に、今後の文字研究に関して蛇足を付け加えさせて頂きたい。その昔(筆者が文字研究に取り組むよりも前), 文字研究はパターン認識研究をリードしていたように思う。郵便区分機は当時の輝かしい成果の一つであろう。リードできていた理由には、文字画像は小さいので限られた計算機リソースでも動かせたことや、ETL データセットの存在、クラスの明確さなど、様々なものがあつたと考えられる。これらの理由により、単に「研究がやりやすかった」からリードできたという見方もあるだろう。しかし筆者は、「文字にしかできない研究, 文字だからやりやすい研究を行っていたから」と捉えたい。すなわち、前述の「研究のやりやすさ」を含めて、文字ならではの性質を適切に活用していたために、リードできていたのだと考える。

今後の文字研究が、パターン認識の一応用分野に留まるのか、それとも「文字にしかできない研究, 文字だからやりやすい研究」を他の分野に先んじて推進していくのか? 本稿からも明らかのように、筆者は後者を目指したい。そのために、文字を単なる画像パターンとして見るのではなく、「文字とは何か」について今後も考えていく所存である。その際に、深層学習を始めとする様々な機械学習や最新の画像情報学技術、そして「対象の性質を見抜く力」が役に立つことは、これからも間違い無いであろう。

謝 辞

最初にパターン認識・メディア理解研究会 (PRMU) にて発表させていただいたのが 1996 年。それから 20 年間、150 件を超える PRMU 発表を行い、自分自身および当研究室の学生諸君が大変お世話になった。文字研究を学んだのも、文字について考える仲間に出会えたのも、この「PRMU という場」であったと思う。二度の PRMU グランドチャレンジ企画 [29], [30] に参画させて頂いたのも、一生の宝と思っている。この PRMU という場を与えてくださった PRMU 関係者の皆様に深く感謝したい。

この度は、「画像および時系列パターンの認識・解析技術の開発と多分野応用」に対して、電子情報通信学会よりフェローの称号を頂いた。未だに薄氷を踏む思いで研究を進める自分がフェロー称号を拝受できたのは、(故) 迫江博昭先生、国内外の大先輩方、協働研究をさせていただいた皆様のご指導が無くては考えられない。そしてもちろん、研究室で共に頑張ってくれた皆様のご尽力が無くては絶対にあり得ない。皆様には心よりお礼申し上げます。今後も引き続きご指導いただければ幸いです。

文 献

- [1] C. von Bearnensquash, "Paper gestalt," In *Secret Proceedings of CVPR*, 2010.
- [2] 小川英光 (編著), パターン認識・メディア理解の新たな展開一

- 挑戦すべき課題一, 電子情報通信学会, 1994,
- [3] 内田誠一, 井手将太, Dipesh Dangol, "文字認識は CNN で終わるのか?", 信学技報, PRMU2015-133.
- [4] R. Gao, S. Eguchi, and S. Uchida, "True color distributions of scene Text and background," *Proc. ICDAR*, 2015.
- [5] R. Gao, et al., "Visual saliency models for text detection in real world," *PLoS ONE*, 9(12), 2014.
- [6] 山本和彦, "[フェロー記念講演] 脳に淘汰された文字とパターン認識技術," 信学技報, PRMU2004-226, 2005.
- [7] P. T. Daniels and W. Bright (Eds), *The world's writing systems*, Oxford, 1996.
- [8] T. Flash and N. Hogan, "The coordination of arm movements: an experimentally confirmed mathematical model," *J. Neurosci.*, 5(7), 1985.
- [9] Y. Wada and M. Kawato, "A theory for cursive handwriting based on the minimization principle," *Biol. Cybern.*, 73(1), 1995.
- [10] R. Plamondon and M. Djioia, "A multi-level representation paradigm for handwriting stroke generation," *Hum. Mov. Sci.*, 25(4 5), 2006.
- [11] ローダ ケロググ, 児童画の発達過程 なぐり描きからピクチュアへ, 黎明書房, 1971.
- [12] デザインノート編集部, ターゲットから発送する文字のデザイン, 誠文堂新光社, 2019.
- [13] A. T. Poffenberger, "A study of the appropriateness of type faces," *J. Applied Psychology*, 7(4), 1923.
- [14] R. C. Davis and H. J. Smith, "Determinants of feeling tone in type faces.," *J. Appl. Psychol.*, 17(6), 1933.
- [15] 小林 章, フォントのふしぎ — ブランドのロゴはなぜ高そうに見えるのか?, 美術出版社, 2011.
- [16] 山王丸 榊, ろごたいぶっ! — マンガ・アニメ・ラノベのロゴを徹底研究する本, 立東舎, 2015.
- [17] Y. Shinahara, T. Karamatsu, D. Harada, K. Yamaguchi and S. Uchida, "Serif or Sans: Visual font analytics on book covers and online advertisements," *Proc. ICDAR*, 2019.
- [18] 杉浦康平, 文字の靈力, 工作舎, 2014.
- [19] MdN 編集部, "特集: キャラの声をもフォントで再現する方法," 月刊 MdN 2017 年 2 月号, エムディエヌコーポレーション, 2017.
- [20] 高田雄吉, ロゴ ロジック, バイインターナショナル, 2013.
- [21] S. Juni and J. S. Gross, "Emotional and persuasive perception of fonts," *Percept. Mot. Skills*, 106(1), 2008.
- [22] 川口維文, 牛久祥孝, 内田誠一, "情景内文字情報を考慮した画像説明文生成," 信学技報, PRMU2017-131, 2018.
- [23] A. F. Biten, et al., "ICDAR 2019 Competition on scene text visual question answering," *Proc. ICDAR*, 2019.
- [24] H. Zhang, et al., "StackGAN: Text to photo-realistic image synthesis with stacked generative adversarial networks," *Proc. ICCV*, 2017
- [25] Ø. D. Trier, A. K. Jain, and T. Taxt, "Feature extraction methods for character recognition-A survey," *Pattern Recognition*, 29(4), 1996.
- [26] S. Ide and S. Uchida, "How Does a CNN Manage Different Printing Types?," *Proc. ICDAR*, 2017.
- [27] 斎藤泰一, 山田博三, 山本和彦, "手書漢字の方向パターン・マッチング法による解析," 信学論 D, J65-D(5), 1982.
- [28] ハンス ロスリング, オーラ ロスリング, アンナ ロスリング ロンランド, *FACTFULNESS —10 の思い込みを乗り越え、データを基に世界を正しく見る習慣*, 日経 BP, 2019.
- [29] 鷺見和彦, 内田誠一, 佐藤真一, 佐藤洋一, 日浦慎作, 福井和広, 馬場口 登 "パターン認識・メディア理解の 10 大チャレンジテーマ (パターン認識・メディア理解のグランドチャレンジ小特集)," 電子情報通信学会誌, 92(8), 2009. (※同号には他にも関連記事あり)
- [30] 内田誠一, 前田英作, "パターン認識・メディア理解の現状, そして Open Idea へ電子情報通信学会誌 (小特集 パターン認識・メディア理解の機能拡張に向けた Open Idea)," 101(10), 2018. (※同号には他にも, 安倍 満, 木村昭悟, 船富卓哉, 中澤篤志, 松下康之, 山崎俊彦による関連記事あり)