研究室紹介

中野研究室の紹介

中野 康明
Yasuaki Nakano
九州産業大学 情報科学部 知能情報学科
Department of Intelligent Informatics, Faculty of Information Science, Kyushu Sangyo University
ysnakano@gakushikai.jp, http://www.is.kyusan-u.ac.jp/~nakano/

1. はじめに

「研究室紹介」を書くことになりましたが、平成 15 年 4 月に赴任してからまだ 2 年しか経っておらず、研究室の成果がこれだというのはあまりありません。

そこで、私が今までどんなことをやってきたか紹介し、今後やりたいことを書きます。

1·1 経　　歴

学歴と職歴をまとめて書きます。

- 昭和 36(1961) 年 3 月: 東京大学工学部応用物理学科卒業
- 昭和 38(1963) 年 3 月: 東京大学大学院数物系研究科応用物理学専攻修了
- 昭和 38(1963) 年 4 月: (株) 日立製作所入社。中央研究所勤務
- 平成元 (1989) 年 3 月: 信州大学教授工学部
- 平成 15(2003) 年 4 月: 九州産業大学情報科学部教授

この表で「信州大学教授工学部」というのを妙な肩書きと感じるかも知れませんが、これが正式です。

2. 日立中研での研究

日立製作所で 26 年間いろいろなことを研究し、その後大学で研究しました。

一般に民間会社では一つのことを長くやるのは難しいのですが、幸か不幸か、どちらかというと同じテーマを長くやっていたように思います。

2·1 東京大学時代

大学と大学院では、数理工学コースに在籍しました。この数理工学コースというのは、日本が第二次世界大戦で敗北し、占領軍 (アメリカ軍) によって航空工学の研究が禁止された時期に、工学部航空工学科の俊英が転進して作った応用数学科の後身です。

日本の独立とともに航空工学科が再建され、多くの先生は応用数学科から航空工学科に戻られましたが、少数の優秀な先生が残ったのが数理工学コースです。学科としては小さ過ぎて応用物理学科の 1 コースになりました。

私の師事した先生は、航空工学ではプロペラの研究で世界的な業績を挙げられましたが、数理工学コースに残って、工学原理を幾何学により統一的に理解することを信条とされていました。

§1 数理音声学

大学での卒業研究は違うのですが、大学院に進学し「数理音声学」をテーマとしました。ちょうどその頃、先生が「人間の音声現象は射影幾何学で理解できる」と発想され、その思想の下に学生達に先生の理論の実験的な裏付けをせよと命じられたのです。

先生の高邁な理論を完全には理解できず、「わしの理論から出る筈の実験データとは逆の結果ではないか。理論を裏付ける実験をして貰わないと困る」と叱られることも度々でしたが、「考えていた理論とは異なるデータだが、こう考え直せば筋の通った説明ができる」と、先生の方で理論展開を変えられることもありました。

そうこうしているうちに修士が終わり、いろいろな事情がありますが、日立製作所中央研究所 (以下では日立中研と略します) に入ることになりました。

2·2 日立中研時代

そのとき日立中研では音響研究を始めるために研究人員を集めていたのですが、研究室長予定者の本心は音声を研究したかったのです。しかし、商売になるステレオ (家電品) を研究の主力とし、音声研究はすぐには商売にならないから、長期的に新入社員にやらせようと考えていました。丁度そこに飛び込んだ話が私だったのです。

日立中研では次のような研究に従事しました。細かいテーマまで含めるともっとあり、かなり省略しています。詳細は

www.is.kyusan-u.ac.jp/~nakano/kenkyu.htm

を参照して下さい。

(1) 音声合成
(2) ソナー信号検出
(3) 音声認識
(4) 印刷漢字認識
(5) 手書き文字認識
(6) 文書理解
(7) 手書き文字認識結果の知識処理

これらを総合してまとめると「パターン情報処理」の研究ということができます。簡単に各項目の説明をします。

§1 アナログ型音声合成

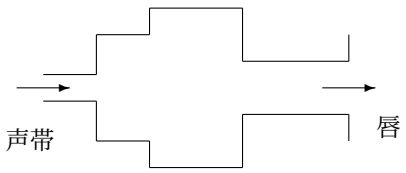
音声合成というのは簡単に言えばコンピュータで音声を発生することです。その原理を大きく二つに分けると「録音編集合成」と「法則合成」です。

録音編集合成というのは、「博多」「行き」「１０時」「３０分」「発列車は」「５」「番線から発車します」というように、音声の要素を録音しておいてつなぎ合わせて文章音声を合成するものです。例えば「博多」のところを「小倉」や「鳥栖」に、「１０時」を「１８時」、などなどと入れ替えれば、多種多様の文章音声を合成することができます。

録音編集合成では録音されていない単語を合成することはできません。法則合成は、任意の文字列からその音声を合成しようとするものです。そのためには音声の発生原理に立ち返って音声波形を合成しないといけません。

日立中研では、音声発生原理として「声道模擬」を採用しようと決めました。「声道」とは声帯から唇までの空洞を指しますが、その中の音波の伝わり方を計算すれば唇から放射される音波が出るというものです。

声道は成人男子では 17cm 程度の長さです。これを等断面積の管の連続と近似すると、例えば図１のような形になります。



一様音響管の連続で近似した声道

図 1 声道模擬での音響特性生成の説明

その音波の伝わり方は微分方程式で記述されますが、それをアナログ計算機で解こうということにしました。当時のディジタル計算機では速度的に実時間で解くことは不可能でした。1 秒の音声波形を計算するのに数十分もかかってしまうからです。アナログ計算機で解くと実時間で解が出ます。唇のところで観測される波形をスピーカーに通せば、音として聞こえるのです。

このアイデアは当時の電気試験所 (現在の産業総合総合研究所) で出され、実現できるという見通しの下で日立に発注されることになり、日立中研でも音声研究の立場から協力することになりました。

§2 ディジタル型音声合成

電気試験所に納入した装置は工場で作りましたから確実に動きました。協力するだけではつまらないから、自分のところでも欲しいと思って私が作りましたが、難産で満足に動きませんでした。

この試作装置はともかく、法則合成型の音声合成は現在でも実用には今一つですから、当時は全く使い物にはなりませんでした。

音声合成で当時実用化されていたのは録音編集型合成装置です。時刻通報とか電話番号自動応答などアナログ式テープレコーダーで小規模には実現されていました。

しかし、合成できる単語数を増やそうと思うとディジタル型になります。日立が国鉄 (現 JR) の列車予約システム用に開発・納入した音声合成装置は、ディジタル型合成で日本最初の大規模なシステムで、世界的にも注目されたものです。

この開発は日立中研と工場が協力して行いました。この研究グループにも参加しましたが、残念ながら私の担当した部分は実用システムに生かされていません。

§3 ソナー信号検出

音響研究室に水中音響情報処理の話が舞い込みました。具体的には潜水艦の所在を音響によって突き止めようというものです。海の中では電波が通りませんから、頼りになるのは音だけなのです。

潜水艦探知を行うシステムをソナーと呼びます。ソナーはアクティブとパッシブに分かれます。

アクティブソナーは電波で言えばレーダーに相当するもので、こちらから音響パルスを送って反射波を捕らえて、近海に潜水艦がいるかを判断します。相手にも自分の所在が判ってしまいますから、使用するにはいろいろな条件があります。

パッシブソナーは相手潜水艦の発生する音を検出し、所在 (とできれば種類) を判定するものです。私の研究したのはこちらです。

技術的には、当時の日本ではアナログ処理が主流であったのに対し、ディジタル信号処理を持ち込んだ研究は早かったと思います。ただ、アメリカではもちろんディジタル信号処理が主流になりかけていましたから、今から思えばそんなに威張れる研究ではありません。

§4 音声認識

音声合成が一段落し、音声認識を手がけました。認識のためには音声分析が必要で、音声合成で提案されていた PARCOR 方式などを調べていました。

そのとき、日本電気の迫江博昭氏 (現九大教授) が提案されたのが、現在では DTW (Digital Time Warping) と呼ばれている方式です。提案されてすぐその有用性は理解できました。

ただちに追試を行い DTW の有効性を確認しました。それだけでは面白くないので、認識に使用する音声の分

析パラメータとして何を使ったら良いか系統的に調べ、その結果を論文にしました。[1]

§5 印刷漢字認識

音声分析を研究している頃、日本で音声・画像・物体・文字などのパターンの認識が研究の花形になってきました。

ちょうどこの頃、通称「パターン大プロ」と呼ばれる国家プロジェクトを通商産業省が企画していました。巨額の資金を投じてパターン認識研究を加速し、この分野で日本を世界の先進国にいしょうという野心的なプロジェクトです。

日立中研も「パターン大プロ」の一角に食い込みたいと希望しました。そこで、パターン認識に関係する研究者を集めて「研究プログラム」(プロジェクトですが、諸般の事情でプログラムと呼びました) を作ることになりました。

私の属する研究室のリーダーがプログラムリーダーに就任し、いろいろな調査を行いました。調査だけでなく実際に手を下して調べてみようということになり、対象として印刷漢字の認識を選びました。

印刷英数字を読む機械は実現されており、手書き数字の読み取りも実現寸前でした。それでも、印刷漢字の読み取りは困難ではあるが挑戦しがいのある目標と思われました。パターン大プロでも研究目標の一つに上がっていました。

そこで、研究プログラムでは何人かの研究者が印刷漢字認識の実験を行い、実現可能性を示そうとしたのです。日立中研には文字認識研究のグループもありましたが、手書き文字認識に手を取られて印刷漢字認識までは手が回らなかったのです。

私があるアイデアで実験してみたら意外にうまく行くことが判りました。この関係でいくつか論文を書き [2]、それらをまとめて学位論文にしました。印刷漢字認識を手がける研究者はほとんどなく、ちょっとした実験をやればすぐ論文になったのです。

日立はパターン大プロに印刷漢字認識と物体認識で参画することができました。しかし、印刷漢字認識には他社も参画していたのに物体認識は日立だけだったという事情もあって、日立は物体認識に集中するため印刷漢字認識からは身を引きました。

§6 手書き文字認識

パターン大プロから撤退したからといって、印刷漢字認識研究を終了したわけではなかったのですが、日立では手書き文字認識が大問題になっていました。

手書き数字認識の巨大マーケットとして、郵便区分機用の郵便番号読み取りがありましたが、日立は経営判断から手を引き、研究はコンピュータ入力装置としての数字認識に集中していました。この経営判断は結果としてはうまく行って、一時は日立の文字認識装置 (OCR*1) は良く売れていました。

ところが、手書き数字から手書き英字に拡張する頃から他社も OCR に力を入れてきました。郵便区分機を郵政省に納入した二つのメーカーが特に強敵ですが、強力な研究所を有する電電公社 (現 NTT) も参入し、手書きカナ認識で激しい競争にさらされました。手書きカナ文字認識の巨大ユーザとして労働省があり、各メーカーに競争させていましたが、結局日立は労働省には納入できませんでした。

この結果、次期 OCR では絶対敗北は許されないとして、文字認識研究は会社上部が監視する「特別研究」に指定されました。研究の主たる目標は手書き英数カナ文字 (ANK) の認識です。研究だけではなく、工場が担当する開発も含めた研究開発全体がプロジェクトになり、全社的に推進することになったのです。

印刷漢字認識がほぼ終結したこともあって、印刷漢字認識グループは手書き文字認識グループと合体し、全面的に手書き ANK 認識に取り組むことになりました。とにかく忙しいし、上層部からは進行状況を監視されるし、で研究開発は大変でした。

私が特別研究にどの程度貢献したかは良く判りませんが、研究結果は工場で製品に組み込まれました。当時工場におられた花野井設計課長 *2をリーダーとする設計グループにより OCR 新シリーズが完成し、他社との激しい競争にさらされながら、誤認識が少ないという定評で売れたようです。

誤認識については、認識手法の評価に "$R+10E$" という尺度を使用しました。ここで "E" は OCR が自信をもって答えを出したのに誤っている率を示し、誤認識率と呼びます。"R" は OCR に自信がなくて「わかりません」と降参する率で、リジェクト率と呼びます。リジェクトは特別なコード (たとえば "?") で出力され、人間が原稿を見て入力します。

それまでの文字認識研究では、認識方式 (OCR) の比較に認識率 "C" を用いていました。認識率とは入力文字に対して正解が出た率ですが、OCR は文字を読む機械ですから "C" を尺度にするのは当然と思われていたのです。

しかし "$C=1-E-R$" ですから、"C" で比較するのは "$R+E$" で比較するのと同じです。ということは、誤認識とリジェクトを同じウェイトで評価していることになります。少し考えれば判るように、誤認識があるとなかなか発見できません。リジェクトの場合は機械に読めなかったことが表示されますから、誤認識よりははるかに望ましいと言えます。

そこで手法の評価に "$R+10E$" を使ったわけです。1 個の誤認識はリジェクト 10 個に相当する悪さだと考えた、とも言い替えられます。"$R+10E$" を小さくするには、多少リジェクトが増えてもとにかく誤認識を減らせ

*1 Optical Character Reader: 光学式文字読取機

*2 情報科学部の花野井助教授です。

ということで、結果的に誤認識を減らすことになったのです。

OCR を製品化した後の話ですが、日本を代表する文字認識研究者が「日立の OCR の誤認識は世界一少ない」と言って下さいました。残念ながら解説などには書いて下さらなかったので宣伝には使えませんでしたが、評判は広まっていたようです。

もう一つ残念なのは、”$R+10E$” で認識方式を評価していることを外部に発表しては困ると言われ、論文などで書けなかったことです。かなり後になって、日本のある国立研究所が ”$R+10E$” で方式評価を行うという考えを論文に書き、世界ではこの研究所が考え出したと思っているようです。

似たような考えを持つ人は世界中に大勢いるので、あるアイデアで結果が出たらすぐ論文を書くことが重要な例の一つです。

§7 手書き文字切り出し

特別研究の次の目標だった小型 OCR の開発の中で、研究自体として意味があり、現在でも有効な手法として、「多重仮説検定法」があります。これは文字切り出しに関係しますが、詳細は次節に述べます。

§8 文書理解

小型 OCR の次の目標では、文字以外に何を認識するかが問われました。答えとして出したのが「文書理解」です。

「文書理解」はかなり大げさな名前で欧米では「文書解析」と呼ぶことが多いようです。誰が言い出したか判りませんが、日本では文書理解 ('Document Understanding') と呼びます。

文字認識とは、一字一字の文字を観測して何という字かを当てることです。たとえば「あ」という字をコンピュータが見ても、これは単なる模様です。コンピュータでこの情報を利用するためにはこれを文字コードの $(2422)_H$ に変換することが必要で、この処理が文字認識です。

実際の文書では、一文字の認識だけではなく紙面全体がどういう構造になっているか解析する必要があります。例えばこのページの第 1–2 行を見て下さい。一文字づつの認識が完全にできたとしても、このページが二段組になっていることが判らずに左から右に読み進めると

ということで、結果的に誤認識を減らすことになったのそれぞれのキャプション*6 があり、です。を持っています。

となって、わけの判らない結果になります。

文書に文字だけしか書いてなく、一段組の単純な構造でも、表題、筆者、所属、章節の見出し、柱[*3]、ノンブル[*4]、ルビ[*5]、など、特殊な構造がかなりあります。

さらに、多くの文書では写真、画像、図、表、数式や、それぞれのキャプション[*6]があり、きわめて複雑な構造を持っています。

これらの文書の構造をレイアウト構造といいます。その他に文書の論理的な関係を示す論理構造もあります。

既存の文書をコンピュータ電子化する場合、単純にスキャンするのではでなく、文書のレイアウト構造を解析し、文字の部分 (テキスト領域と呼びます) を文字認識することが好ましいといえます。

このような処理を文書理解と呼びます。私が研究を始める前から文書理解は内外で研究されていましたが、私の知る限り画像処理手法によって画像の特徴から文書画像を分解する方法がほとんどでした。

私は文書には文字が含まれる以上、完全に文書を理解するためには文字を認識することが必須であると考え、その方向で研究しました。

全ての形式の文書とは言いませんが、事務処理で重要な表形式の文書について、文字認識と協調した文書理解手法を提案し、実験システムで効果を示しました。[3]

これとは別に全く新しい考え方として、文書に関する知識を書式として記述し、文書画像と書式とを照合して構造を理解する手法も提案しました。世界的にも新しい考え方で論文も書きましたが、その後発展していません。[4]

§9 手書き文字認識結果の知識処理

経理伝票の数字の認識などでは、対象はデータですから一字の認識結果が誤っていては役に立ちません。

しかし、漢字の認識の場合、文字は文章の構成要素として使用されることが多いので、言語的な知識を使うと個々の文字の誤認識があっても修正可能なことがあります。

たとえば、都道府県の記入の場合、「鹿児島」の 3 文字目を誤認識して「鹿児鳥」と出力しても、住所の知識を使って「鹿児島」と修正することができます。

ただし、度を過ぎると弊害もあります。某メーカーのデモで、OCR が「信州大学」を正しく「信州大学」と認識したのに、知識を使って「九州大学」に修正してしまい、信州大学関係者から叱られたという話が伝わっています[*7]。

印刷漢字認識のときに、既に知識処理の重要性に気づいており、多少の研究もしましたが [5]、本格的に必要になったのは手書き漢字認識を始めてからです。ただ、この研究を行っている途中で信州大学に転職しましたので、自分では最後まではやっていません。

3. 多重仮説検証法

もう 20 年近く前になりますが、「多重仮説検証法」は私の一つの重要な成果です。私が全部を一人でやったの

*3 ページ最上部にある見出しを「柱」といいます。
*4 ページ番号を「ノンブル」といいます。
*5 振り仮名を「ルビ」といいます。
*6 図などについている表題や説明文を「キャプション」といいます。
*7 限りなく実話に近いフィクションです。

ではなく共同研究の結果ですが、メーカーに限らず現在の研究では共同作業が非常に重要になっています。

それまでの手書き文字認識では、文字を別々の枠に記入することにしていました。それでは帳票の利用率が悪いので、OCR 利用者から数文字を一つの枠 (フィールド枠) に記入したいという要求が強かったのです。

しかし、フィールド枠に複数文字を記入させると文字が接触してしまい、認識できないという問題もあります。したがって、接触文字の分離が重要な問題になります。

3·1 接触文字の分離

接触文字の分離ではいくつかの問題があります。

- 妥当な分割点候補の決定
- 複数の分割点候補からの最適点の決定

§1 輪郭解析による分割点候補の決定

従来からも接触文字の分割は試みられていなかったわけではありません。しかし、その方法は「周辺分布」を利用し、その谷の部分で分割するというものでした。

われわれは周辺分布を採用しませんでした。そもそも、周辺分布は私が印刷漢字認識で用いてある程度成功した方法 [2] で、それを使わないというのはおかしいと思われるかも知れませんが、周辺分布の欠点が判っていたから使わなかったと言えるかも知れません。

共同研究者が、文字接触箇所の分析により輪郭解析手法が有効であろうと提案しました。輪郭解析により接触箇所が推定できれば、分離も輪郭解析により比較的容易にできます。

図 2 で、最上行の入力で接触している箇所を切断した結果が第 3–4 行で示される分割候補ですが、かなり自然な分割が得られていることが判るでしょう。

§2 多重仮説検定法

ところで問題なのは輪郭解析によって得られる分割候補が複数ある場合です。分割するかしないか、も問題になります。図 2 にその事情が示してあります。ある箇所で切断できるというのは仮説に過ぎませんから分割しない場合も含めて「複数の切断仮説が立つ」と呼びます。さらに分割した文字成分[*8]をどう組み合わせて一つの文字と見るかによって、数多くの切断仮説が立ちます。

文書画像の中から一文字づつを分離して認識部に送る部分を「切り出し部」と呼びます。それまでの切り出し部では、文字の切り出し方はたった一通りでしたから、分離した文字パターンを単純に認識部に送るだけで悩む必要はなかったのです。しかし、われわれの手法では複数の切断仮説が立ちますから、どれか一つに決定するとなるとどう決めたら良いかが問題になります。しかし、それらの切断仮説のどれが正しいかは切り出し部では決定できません。切り出し部には「知能」がないからです。

私たちが出した結論は「切り出し部では複数の仮説を立て、そのどれが正しいかは認識部で決める」というものでした。結論の先送りといってもいいかも知れません。これを多重仮説検定法と名づけました。

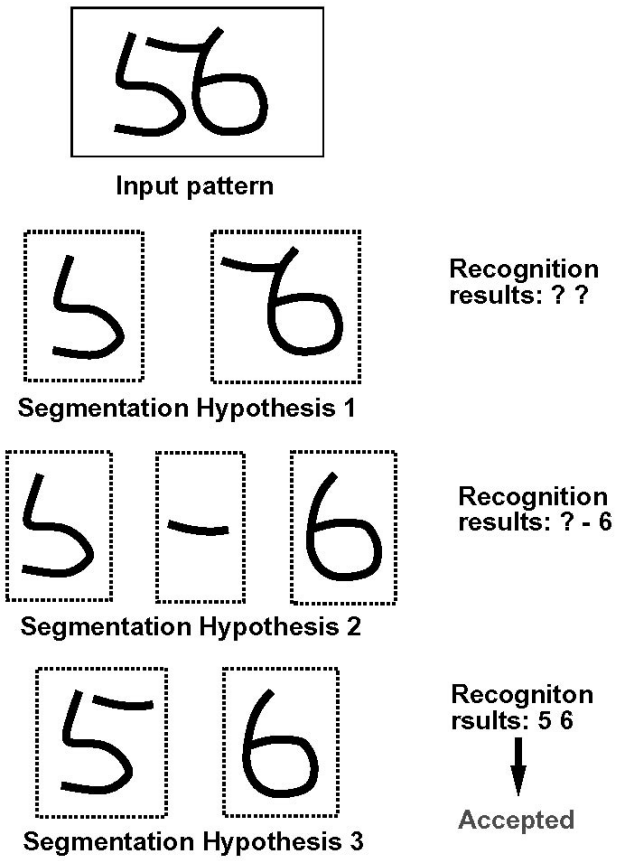


図 2 多重仮説検定法

図 2 に仮想例を示します。最上行が入力された画像 (文字パターン) です。第 2–4 行が切り出し仮説とその認識結果を示します。

三つの仮説が立ちますが、そのうちの二つ (第 2–3 行) は認識結果が出ません。つまり「そのように切り出しても文字として読めません」と言っているのです。このように、認識部でリジェクトになる (認識結果が出ない) 仮説は捨てます。図 2 第 4 行のように、全ての文字パターンが認識された場合[*9]、その仮説を採用することにしました。

その効果は絶大でした。多重仮説検定法により接触文字の認識の問題が少なくとも手書き数字については解決されました。[6]

§3 Oversegmentation

ところで、多重仮説検定法を外国では 'oversegmentation' と言います。私が 10 年ほど前にアメリカのいろいろな研究所で多重仮説検定法を説明したところ、「oversegmentation についてはつい最近議論した」という反応が多かったのには驚きでした。なんで驚いたかというと、多重仮説検定法はさらにその 10 年ほど前に提案したもので、今さ

*8 分割する前から離れている場合もあります。

*9 認識部が誤ると危険です。しかし、私達の数字認識方式は誤認識が少なかったので踏み切りました。

ら議論するほどの新規なアイデアとは思っていなかったからです。これも残念な例ですが、われわれの名付けた「多重仮説検定法」を英語で宣伝しなかったため、別の名で呼ばれてしまいました。

欧米では手書き文字は単語として筆記しますから、個別文字として認識する前に連続筆記体単語から文字を分離しなくてはなりません。ところが分離候補点は複数ありますから、必然的に多重仮説検定法を使用せざるを得ず、多くの研究所で研究を始めたわけです。

私が連続筆記体英単語の認識を実験した結果を図 3 に示します。細かい説明は省略しますが 'Tokyo' という連続単語を文字候補に分断し、それを組み合わせた多重仮説を作って認識部で確定しています。[7]

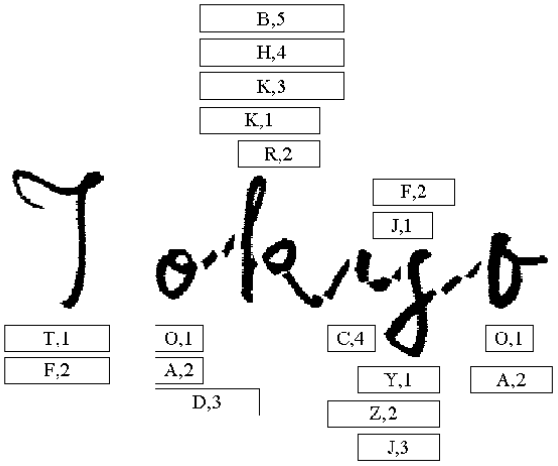


図 3 英単語認識の例: 上下の

ただし、誤解されないように付言しますと、この論文が世界で初めての連続筆記体英単語の認識ではなく、アメリカやフランスで先に研究成果が出ており、連続筆記体英単語に限って言えば後追いに近いものです。私たちの研究は単語の分割方法と対象の選び方 (日本の都市名) に新しさがあり、進歩という観点では大きなものではありません。

3·2 漢字への応用

これまでの説明では省略しましたが、多重仮説検定法は接触文字だけを考慮したものではありません。複数の文字成分の統合方法としても重要です。

分離した成分が文字として認識できたかの情報を利用することも多重仮説検定法に含まれます。漢字のように多数の文字成分が組み合わさって構成される場合には本質的な手法と言えます。

欧米の文字では、一つの文字が分離した成分を持つことは稀です。しかし、漢字は偏 (へん) や旁 (つくり) などを組み合わせて作られていますから、偏や旁が別の文字になる場合が問題になります。

例を挙げればきりがありませんが、「鶏」という画像を認識したいとします。この画像は「奚」と「鳥」という二つのパターンに分離できますが、第一成分は文字としては認識できません[*10]。したがって、この画像は分離してはダメで一括して「鶏」と認識した方が良いと結論します。

§1 言語処理の利用

それでは、文字認識結果を利用して多重切り出し仮説の中から最適組み合わせを選んでよいのでしょうか。

そうではない、という例を図 4 に示します。

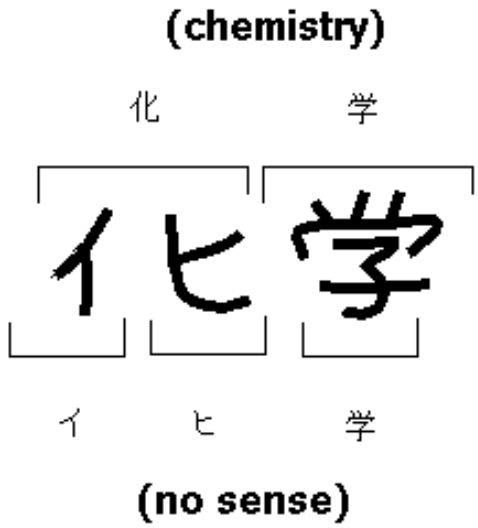


図 4 言語処理が必要な例

「化学」という画像は、文字のレベルでは「イヒ学」とも「化学」とも読めます。しかし、「イヒ学」という単語はありませんから、そのような切り出し仮説は否定され、「化学」が候補として確定します。

この手法については日立中研在職時に特許を出願しました。最近の郵便区分機では手書きされた漢字住所の読み取りを実現していますが、そのためにはこの手法は必須だと思います。しかし、出願が早過ぎたのか、やっと実現された頃にはそろそろ期限切れです。日本の文字認識研究者の世界では「特許は期限が切れた頃に使われる」というジンクスがあります。

ところで、上記の「化学」の例では単語辞書と照合すれば正しい仮説が選択できます。しかし、数年前に単語照合だけでは片付かない例があることに気付きました。

OCR ではありませんが、テレビのニュース番組で[*11]ある美人アナウンサーが「旧中山道 (きゅうなかせんどう)」を「いちにちじゅうやまみち (1 日中山道)」と読んだという話があります。

---

*10 これは常用漢字の範囲での説明例で「奚」は文字です。

*11 本学部の某先生がたまたま番組を見ておられ、ニュースではなくバラエティー番組であり、実際の内容も少し違うそうです。

この原稿を OCR が読んだとしたらどうなるか面白い問題です。「旧中山道」はちょっとした辞書にありますが[*12]、「１日中」＋「山道」も複合語として成立します。したがって、単語照合だけではどちらが正しい解釈かは確定できません。

この例では、それまでの話の流れを知らないと正しく読めないわけで、自然言語処理の分野で「談話理解」と呼ぶ高度な処理が必要になってきます。

4. 大学での研究

信州大学に移ってからも日立中研での研究を基本的には続けましたが、学科主任の先生などから人工知能関係の研究も加速して欲しいとの要請があったこともあって、自然言語処理や意味ネットワーク、機械翻訳なども手がけて見ましたが、この分野ではあまり成果が出ていません。

関連する分野でマルティメディア研究にも手を出しましたが、研究の中心はやはりパターン認識関係でした。パターン認識と人工知能との接点といえるかどうか、ニューラルネットワークも流行に乗ってやりました。

九州産業大学に赴任してからも信州大学での研究の延長上で進めています。

3 章で述べた手書き英単語認識の研究は信州大学で行ったものですが、便宜上 3 章で書きました。

大学に移ってから行った研究で多少とも成果の出たものでは次のものがあります。

(1) 複数の OCR の結果の統合
(2) 顔画像の抽出と認識
(3) 情景からの文字の抽出と認識
(4) オンライン文字データベースを利用したオフライン文字認識の高度化

ここでは、最初の項目についてだけ説明します。

4･1 複数の OCR の結果の統合

文字認識結果の知識処理については 3･2 節で少し述べましたが、別の形式の知識処理があります。それは、複数の OCR の認識結果を統合して認識性能を上げるというものです。

わかりやすい例で言えば、三つの OCR があって、それぞれの性能は同等とします。同じ文字パターンに対しての出力が "A, A, A" であれば入力は "A" らしいし、"A, H, R" だと何が答えか判らないでしょう。OCR は全てが同じように間違うのではなく、間違い方の癖があります。これを利用しようというものです。

このことを世界で最初に提案した論文ではありませんが、わかりやすい例なので郵政研究所の結果 [9] を用いて実際に見てみましょう。

4･2 満場一致原理

図 5 では、同じ文字パターンセット (10,000 個) に対する三つの OCR: α、β、γ の誤認識の重なり具合を示してあります。誤認識は、α では 36 個、β では 52 個、γ では 41 個です。

この図 5 で、重なり合う領域が同じ誤認識をした文字パターンを示し、数字はその個数です。例えば、α と β が同じように誤った個数は "5＋3＝8" です。三つの OCR が全部同じ誤りをした個数は僅か 3 個であることがわかります。

したがって、もし三つの OCR の出力が「満場一致」であるときにのみ答えを出すとすれば、図 5 の網掛けの部分のように誤認識は僅か 3 個に減ります。誤認識率にして 0.03% にしか過ぎず、個々の OCR より一桁も下がります。

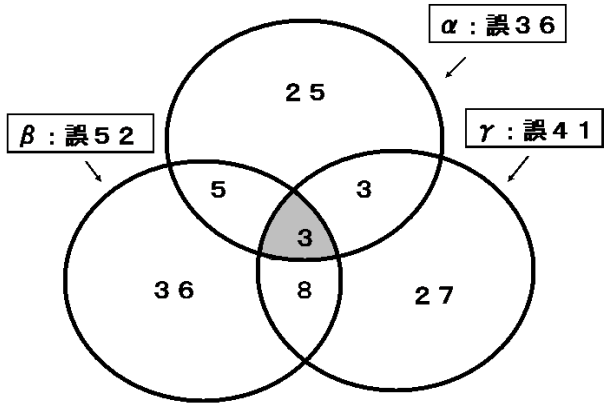


図 5 満場一致の説明

しかし、満場一致の場合は一つでも違う答えを出すとリジェクトになるので、リジェクト数は 104 個も増えてしまいます (図 5 の網掛け部以外の合計です)。

ただし、上に述べた "$R+10E$" 評価を使えば、圧倒的に改良されていることは確認してみて下さい。

4･3 多数決原理

"$R+10E$" 評価では大幅に改良されるとはいえ、リジェクトが大幅に増えるのはいやだ、という考え方もあるでしょう。そのときに有効なのが「多数決」です。

この場合、使用する OCR は 3 個なので、多数決すなわち 2 個以上が同じ答えを出せば、結果として採用しようというものです。

多数決を取ると、誤認識は二つ以上の OCR が同じ誤りをした場合ですから、図 6 の網掛け部のように 19 個に減ります。満場一致には劣りますが、しかし、リジェクトはほとんど出ません[*13]。

---

*12 美人アナウンサーは頭の中の辞書がたいへん薄いとひやかされたわけです。

*13 多数決によってリジェクトが出ないとは言えませんが、非常に少ないでしょう。

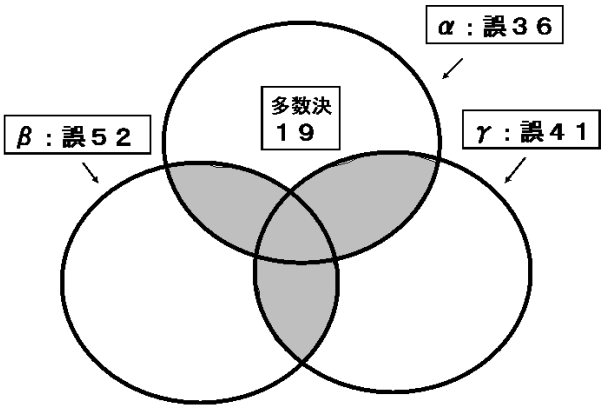


図 6 多数決の説明

したがって、多数決はリジェクトをほとんど増すことなく、誤認識をかなり減らす方式だと言えます。

§1 多数決の歴史

多数決のアイデアはかなり古いのです。日本で 1964 年に特許出願がなされています。[9]

しかし、いくらアイデアがあったとしても 1964 年に実現することは不可能でした。そのころの OCR はハードウェアで実現されていましたから、三つの OCR を組み合わせるとコストは三倍になってしまうからです。

このアイデアが注目されるようになったのは、ソフトウェア OCR といって、文字認識処理をマイクロコンピュータのソフトウェアでやってしまう OCR の実現性が見えてきた頃です。

1992 年にアメリカで 7 種類の OCR を用いて英語印刷文書の認識実験を行いましたが、そのとき多数決を取ると画期的に認識率が上がることを示したのです。この実験自体はハードウェア OCR を用いましたが、ソフトウェア OCR の出現は予想されていました。ソフトウェアなら、何個使用しようが、コスト上昇は知れています。速度は落ちますが、数年立てばコンピュータの速度は一桁上がりますから問題になりません。

したがって、複数のソフトウェア OCR を用いて認識率を上げる研究が盛んに行われ、英語印刷文書や手書き数字について多くの発表がありました。

4･4 印刷日本語文書への適用

印刷日本語文書に使えば、英語と同じように効果があることは予想されます。しかし、当り前だと思ったのかどうか、実験で確認した人はいませんでした。そこで、前勤務先の信州大学で学生の卒業研究としてやって貰い、1998 年の国際会議で結果を発表しました。

使用したのは 6 種類の日本語 OCR で、実験を行った 1997 年当時は全て最新版でした。

対象とした文書は JEIDA'93 データベース [12] と呼ばれるものから選びました。JEIDA とは (社) 日本電子産業協会の英文名の略称です。JEIDA'93 データベースは文書理解のための標準文書画像として使用するため、JEIDA の一組織である「認識型入力方式調査委員会」が集め、公表したものです。

表 1 印刷日本語文書での多数決の効果

OCR	誤り	消失	湧出	認識率 (%)
α	541	77	6	97.2
β	1,349	103	46	93.3
γ	442	31	23	97.8
δ	453	48	13	97.7
ε	413	17	11	98.0
ζ	373	124	21	97.7
多数決	112	15	0	99.3

しかし、JEIDA'93 データベースを用いた研究はほとんど発表されていません。このデータベース収集に関わった一人として、研究に使えるものだということを示すこともこの発表の目的でした。

実験に使用したのは、5 種類の文書の合計 25 ページで、その中の文字数は合計 22,364 文字でした。

6 種類の OCR の誤認識数と認識率を表 1 に示します。誤認識には、通常の誤り (他の文字に誤る) の他に消失と湧出があり、表 1 では別々に個数を示してあります。認識率はこれら全ての誤りの合計を全文字数から引いたもの (正解) の割合です。

英語で効果が確認されているから日本語で追試する必要はない、という考え方が正しくないことがこの結果からもわかります。

日本語では漢字が多く出現します。ところが 3･2 節の知識処理で述べたように、漢字は偏 (へん) や旁 (つくり) があるため、一つの文字を部分に分割して 2 文字にしてしまうことがあります。見かけ上は文字数が増えて「湧出」になります。

逆にいわゆる半角文字を一つの文字として認識することもあります。このときは文字数が減って「消失」になります。湧出と消失が同一行で同時に起きることもあり、このときは文字数が変わりませんから厄介です。

いずれにせよ、湧出や消失が起きると、いくつかの OCR の結果を統合するといっても関係ない位置の文字を対応させることになるから無意味です。次の例は実際に三つの OCR で出力されたものですが、同じ文字が別の位置に現れていることがわかるでしょう。

しよう。★Ｒ８：１２の・Δと１２を
しよう。２８＝１２の，８とＩ２を
しよう。★２８：１２の，８と１２を

図 7 文字数が異なる OCR 出力

ここでは三つの出力しか示していませんが、実際には 6 種類の出力があります。上の例のようにごく短い出力結果でも、詳細に見ると誤認識 (「・Δ」その他) が沢山ありますが、多数決によって多くは修正できるので、いかに文字を対応付けるかが重要です。

文字数が異なる文字列の間で最適照合を得る方法は別の分野で知られていたので、その方法を適用してから多数決を取ることによって表 1 の結果が得られました。*14

4･5 それだけでは解決にならない

4･3 節や 4･4 節で示された認識率の値は優れたものですから、OCR に利用すれば良いと思われるかも知れませんが、実はただちには実用化できません。

説明を省略しましたが、4･3 節や 4･4 節の実験では、OCR で読み取るテキスト領域は作業者がマウスで指定していたのです。複数の OCR でそんな作業を行うのは実際には面倒で使えません。

現在の印刷文書 OCR では文書理解の能力があり、認識すべきテキスト領域を自動的に抽出するようになっています。

しかし、文書理解の性能は各社の OCR で異なっていて、同じテキスト領域が同じように抽出されるとは限りません。図 8 に判りやすい例を示します*15。

図 8 で、左右は同一の文書画像ですが、二つのの OCR で異なった領域抽出が行われています。四角で囲われている部分が一つのテキスト領域で、これらが領域単位で OCR の文字認識部に送られます。

二つのの OCR で同じ領域番号のテキスト領域を対応させたら、前節で説明した方法では全く意味のない結果しか得られません。

文字と同じように、領域が消失したり湧出したりする問題もあります。領域の間の区切りが異なることもあります。

4･6 領 域 対 応

この問題に対して、領域の最適な対応付けを行う方法を考え、やはり信州大学の博士前期課程の学生に修士研究としてやって貰いました。九州産業大学に赴任してから、この結果を拡張して論文にしました。[10][11]

ごく簡単にアイデアを述べますと、二つの OCR の間で領域同士の対応付けを行うものです。領域ごとの対応付けは領域の類似性に基づいて行いますが、領域の類似性はその中の文字行の類似性の総和で、文字行の類似性は出力文字列の一致度に基づいて計算します。

したがって、二つの OCR で抽出された領域の個数をそれぞれ M、N としますと、$M \times N$ の組み合わせについて領域の類似度を計算する必要があります。その中で最も良い類似度を与える組み合わせによって対応付けを決定します。

OCR が K 個あるとしますと、二つの OCR の組み合わせは ${}_KC_2 = \frac{K(K-1)}{2}$ 通りありますから、それぞれについて上記の対応付けを行います。

このように対応付けを行ってから 4･4 節の手法を適用します。厳密にはもう少し入り組んでいますが詳細は省略します [11]。

この手法を 5 種類の OCR の出力結果に適用しました。これらは 4･4 節で用いたもののサブセットです。

使った文書画像は JEIDA'93 データベース 77 ページ、学術論文 50 ページの合計 127 ページです。

表 2 は、各 OCR で領域抽出が失敗した文書数を示します。失敗したかどうかは原文書との目視比較で決定しました。最も成績の良い A でも 18 ページの文書で抽出が失敗しています。最も成績の悪い D では 1/3 以上の文書で失敗しています。

それに対し、提案した自動対応付けでは領域抽出失敗数が 12 ページに減少しました。失敗数が減るのは、抽出した領域が他の OCR とどうやっても対応付けられないような OCR を除外するからです。

自動対応付けに成功した文書で、4･4 節で説明した手法を適用することにより、個別 OCR より文字認識性能が向上することが示されました。除外される OCR があるので多数決の効果は若干薄れます。

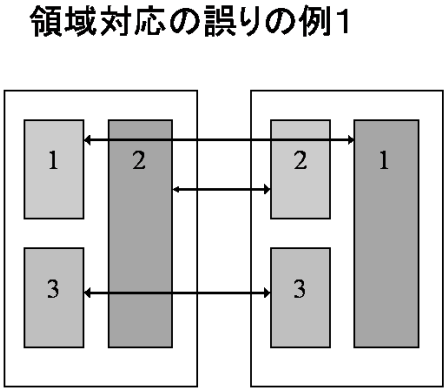


図 8 領域対応付けの誤り (1)

表 2 日本語 OCR で領域抽出が失敗した文書数

OCR	JEIDA	論文	合計
A	14	4	18
B	22	4	26
C	8	10	18
D	24	20	44
E	13	6	19
自動対応	10	2	12

*14 その他、ルビ行が行ごと消えてしまう問題などもありますが詳細は省略します。

*15 これは実例ではなく説明用の仮想例です。

4·7 提案手法の評価

提案した手法で領域抽出が画期的に改善されたかというと、表 2 の結果は満足すべきものとは言えず、個別文字認識で示され劇的な性能改良には遠く及びません。

その理由は、適用できる対象が図 8 のように領域抽出はほぼ同一で順序付けだけが異なっている場合*16に限定されるからです。

さらに、図 9 のように領域抽出自体が異なっているときには適用できません。もう一歩の技術革新が必要なようです。

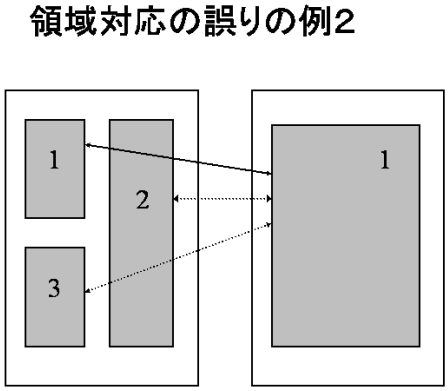


図 9 領域抽出の相違

また、OCR の性能向上はそれぞれのメーカーが行えば良いことであって、本研究のような他力本願的なアプローチに意味があるのか、疑問がないわけではありません。

5. 現在興味を持っていること

文字認識や文書理解の改良も地道に研究して行きますが、パターン認識の手法を一ひねりすると新しい応用が開けることも考えられ、その方向で検討しているところです。

未完成というか、全く手をつけていない状況で詳しく説明するとまずいので、成果が出てから紹介します。それまでのお楽しみにして下さい。

自分でそんな分野の研究に是非参加したいという学生がいたら、研究室に来て下さい。ただし、青春を棒に振る結果になっても知りません。

◇ 参 考 文 献 ◇

[1] Akira Ichikawa, Yasuaki Nakano and Kazuo Nakata. *Evaluation of Various Parameter Sets in Spoken Digits Recognition*. Transactions of the IEEE on Audio, AU–21, 202–209. (1973).

[2] 中野康明 中田和男 中島晃. 周辺分布とそのスペクトルによる漢字認識の改良. 電子通信学会論文誌, 57-D(1):15–22. (1974 年 1 月).

[3] 中野康明 藤澤浩道 国崎修 岡田邦弘 花野井歳弘. 文字認識と協調した表形式文書の理解. 電子通信学会論文誌, J69-D(3):400–409. (1986 年 3 月).

[4] 東野純一 藤澤浩道 中野康明 江尻正員. 書式定義言語を用いた文書画像の理解. 画像電子学会誌, 17(5):267–276. (1988 年 5 月).

[5] Hiromichi Fujisawa, Yasuaki Nakano, Yoshiaki Kitazume and Michio Yasuda. *Development of a Kanji OCR: An Optical Chinese Character Reader*. Proceedings of International Joint Conference on Pattern Recognition, 119–121. (1978).

[6] Hiromichi Fujisawa, Yasuaki Nakano and Kiyomichi Kurino. *Segmentation Methods for Character Recognition*. Proceedings of the IEEE, 80(7), 1079–1092. (1992).

[7] Hirobumi Yamada and Yasuaki Nakano. *Cursive Handwritten Word Recognition Using Multiple Segmentation Determined by Contour Analysis*. IEICE Transaction Information & Systems, E79-D, 464–470 (1996).

[8] Toshihiro Matsui, Ikuo Yamashita and Toru Wakahara. *The Results of the First IPTP Character Recognition Competition and Studies on Multi–Expert Recognition for Handwritten Numerals*. IEICE Transaction Information & Systems, E77-D(7), 801-809. (1994).

[9] 秋元喜一郎 (沖電気工業株式会社). 文字、記号の識別方式. 特許広告公報, 昭 39–017007. (1964 年 6 月 1 日出願).

[10] Hidetoshi Miyao, Yasuaki Nakano, Atsuhiko Tani, Hidesato Tabaru and Toshihiko Hananoi. *Printed Japanese Character Recognition Using Multiple Commercial OCRs*. Journal of Advanced Computational Intelligence, 8(1), 200–207. (2004).

[11] Yasuaki Nakano, Toshihiko Hananoi, Hidetoshi Miyao, and Ken–ichi Maruyama. *A Document Analysis System Based on Text Line Matching of Multiple OCR Outputs*. Proceedings of Sixth IAPR Workshop on Document Analysis Systems (DAS'2004), 463–471. (2004).

[12] http://www.is.aist.go.jp/etlcdb/docidb/. (1993).

*16 多数の OCR でこの条件が満足されていれば良く、全部の OCR でそうなっていなければならないということはありません。