

## 手書き文字時系列筆跡パターン解析 1

4G-7

東山孝生

山中由紀子

中川正樹

東京農工大学

## 1.はじめに

これまで、オンライン手書き文字パターン認識のためにさまざまな手法が試みられてきた。しかし、研究の基盤となるべき共通に利用できるオンライン文字パターンデータベースは存在しなかった。

認識技術の着実な進展のためには共同利用可能なデータベースは不可欠である。我々は今後のペン入力の実用場面を想定してそれに合ったデータベースを作成してきた[1]。このデータベースは1人分約12,000文字から構成される手書き文字時系列筆跡パターンで現在までに当研究室で約30人分、他の協力機関より約70人分が収集されている。データベースに筆記する文字は約10,000文字の文章と約2,000文字の文字列から構成されている。筆記する字体には一切制限を設けていない。また、筆記終了後に誤字脱字を検査し、見つかった誤字脱字で筆記者本人が認めたものを書き直してもらうことでデータベースとしての信頼性を高めている。

本稿では、当研究室で収集した30人分のデータベースを用いて、同一筆記者の筆跡パターンの変動解析についての報告を行う[2][3]。

## 2.同一筆記者の筆跡パターン変動

## 2.1.検査方法

Analysis into pen-trace patterns of handwritten characters  
part 1

Takao Higashiyama, Yukiko Yamanaka, Masaki Nakagawa.

Tokyo University of Agriculture and Technology

データベースの文章部分は平仮名や句読点など繰り返し出現する文字カテゴリを含んでいる。出現回数別のカテゴリ数を表1に示す。1人分のデータベースに2回以上出現する857カテゴリ3239文字を用いて筆跡パターンの変動を調べる。

表1 出現回数別カテゴリ数

出現回数	1	2~5	~10	~20	~100	101~
カテゴリ数	680	554	181	56	49	17

筆跡パターンに変動が起きていることを当研究室で開発された認識システム[4][5][6]を用いることによって調べる。繰り返し出現するカテゴリの中で最初に出現した筆跡パターンを辞書パターンとして学習し、同じカテゴリの他の筆跡パターンとマッチングを行いその類似度を得る(図1)。実際には857文字を学習して、2435文字に対して認識を行う。筆跡パターンが安定しており、変動が少ない場合には得られる類似度が高く、変動が大きい場合には低いと考えられる。なお、類似度は0~1000の値で得られる。

しかし、この手法だけではどのような変動が起きているかはわからない。どのような変動が起きているかを知るためには類似度の低いパターンを目視することが必要である。ただし、あるカテゴリ内の最初に出現した筆跡パターンが、その他の筆跡パターンと比較して大きく変動しているパターンである場合は、類似度がすべて低くなってしまふことが考えられる。

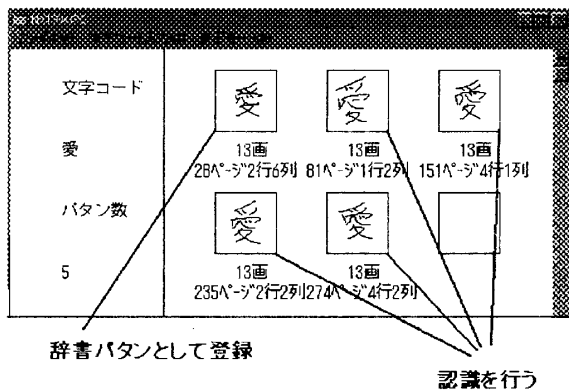


図 1 筆跡パターン変動の検査

### 2.2.変動の個人差

解析の結果筆記者間にかんがりの変動の差が見られた。図 2 に特徴的な 3 人分の筆跡パターン変動検査の結果を示す。nk19 は非常に変動の少ない筆跡パターンセットで 2435 パターン中 1970 パターン (約 81%) が 900 以上の類似度を持っている。それに対して nk14 は変動の多い筆跡パターンセットである。このように筆記者間によって筆跡パターンの変動には差が見られる。

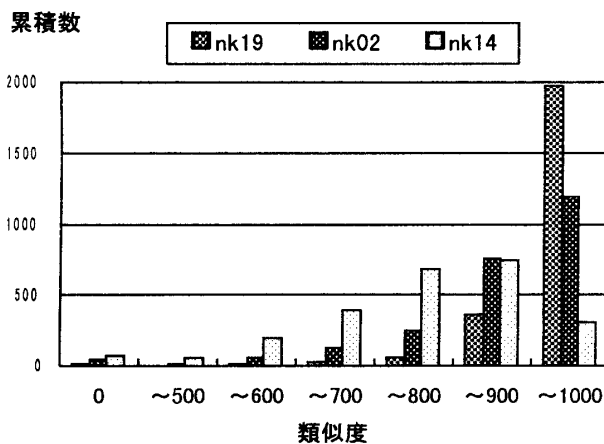


図 2 筆跡パターン変動の個人差

### 2.3.筆跡パターン変動の例

表 2 にデータセット nk11 の「第」を学習、マッチングした結果を示す。この類字度は大き

く分けて 800 台のものと 500 台のものがある。実際に筆記されたパターンを図 3 に示す。類字度が 800 台のパターンは最初のパターンと同じく「第」と筆記されているが、類字度が 500 台のパターンは略字で筆記されていることがわかる。

表 2 マッチングの結果

筆記パターン	2	3	4	5	6	7	8
類字度	849	556	832	595	587	576	595

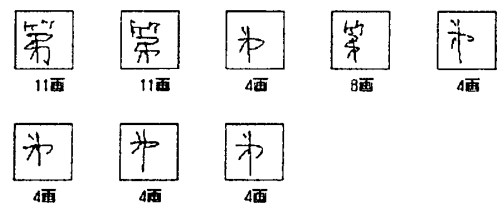


図 3 字体変動の例

### 3.おわりに

本稿では同一筆記者の筆跡パターン変動について述べた。今回検査した多くの筆記者については大きな筆記パターンの変動は見られず、比較的安定していたといえる。逆に一部の筆記者では激しい変動が起こることという結果も得られた。

### 参考文献

- [1] 中川、東山、山中、澤田、レー、秋山：文章形式字体制限なしオンライン手書き文字パターンの収集と利用、信学技報 PRU95-110、43-46(1995.9).
- [2] 山中、東山、澤田、中川：オンライン手書き筆跡パターンの収集とその一解析、情処人文科学とコンピュータ研資 28-1、1-6(1995.11).
- [3] 東山、山中、澤田、中川：手書き文字時系列筆跡パターンの一解析と今後の計画、公開シンポジウム「人文科学とデータベース」講演予稿集、37-42(1995.12).
- [4] 秋山、中川：ストロークのつながりに寛容なオンライン手書き文字認識、画像の認識・理解シンポジウム (MIRU'94) I、67-74(1994.7).
- [5] M.Nakagawa and K.Akiyama:A Lineart-ime Elastic Matching for Stroke Number Free Recognition of On-line Handwritten Characters, Proc. 4th IWFHR, 48-56(1994.12).
- [6] レー、秋山、中川：ストローク数非依存の高速オンライン手書き文字認識手法、情処第 50 回全大、2-61/62(1995.3).