

原 著

パーソナルコンピュータによる筆跡の識別 (II)

若原克文\* 川村 司\*\* 三井利幸\*

IDENTIFICATION OF HANDWRITING BY  
PERSONAL COMPUTER (II)

Katsufumi WAKAHARA, Tsukasa KAWAMURA and Toshiyuki MITSUI

If the following passages, we report the conclusion to try to identify the writer through the singularity of his handwriting by using the theories of factor analysis, discriminant analysis and quantification analysis. First we are going to explain as to how to analyse the Chinese letters which have many strokes. According to our analysis, we could get the following data that we could classify the writer accurately enough to the 1st factor score and discriminant coefficient, which we could get by making factor analysis and discriminant analysis. But in case we handle very simple Chinese letters with a few strokes, we have to judge and decide the identification of the writer subjectively, because we can not get the sufficient data by using the above means. So we made a new trial to get more accurate results to judge by way of the additional use of quantification I and II. We made another analysis and found out the following. It is possible to get enumerative data of the singularity of a written by checking the parts roughly written by a certain writer by using quantification analysis IV. As a result we might say that we could find out the best way to identify the writer by analysing his letter through the above theories, cluster analysis and principal component analysis (which we reported in the latest backnumber). Now we are going to explain the theories and way by using the actual material.

緒 言

前回は報告した主成分分析<sup>1)</sup>による筆跡の識別に続き、今回は同一筆者によって同一時に記載された10文字(以下筆跡)を用いて、因子分析<sup>2)</sup>、判別分析<sup>3)</sup>、数量化理論第II、III類<sup>3)</sup>から筆者識別を、さらに数量化理論第IV類<sup>4)</sup>から文字記載時における筆者の個人内特徴を抽出する方法について検討した。さらにこれらの方法に前回報告した主成分分析を加え、実際の検査試料について筆者

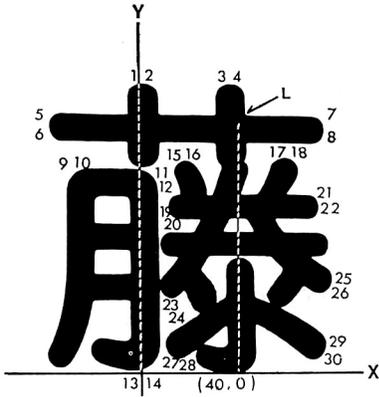
の異同識別を行う方法についても検討した。なお筆跡をコンピュータで処理するための前処理の方法等については前回と同様な方法で行った。また使用したパーソナルコンピュータも前回と同様 NEC9801F を使用した。

1. クラスター分析

前回と同一の方法で筆跡の各座標点を数値化したものを用い、互いに類似したグループごとに分類していく方法であるクラスター分析をワード法で行った。今回は検査文字として「藤」を用いた。図1に「藤」の要素番

\* 愛知県科学捜査研究所  
Criminal Science Laboratory, Aichi Police H. Q.,

\*\* 愛知県立旭丘高校  
Asahigaoka High School, Aichi Prefecture

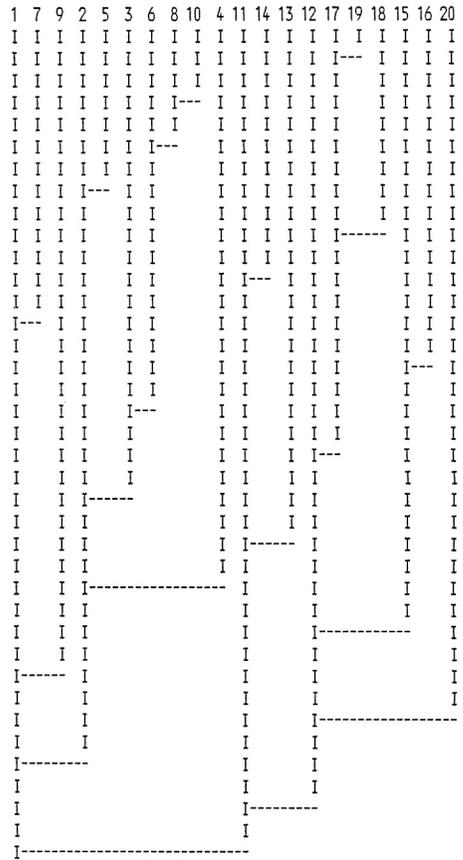


基線は ( ) 座標点とLとを結んだ長さで Y軸に平行  
 図 1 「藤」における基線と要素番号 (座標点) のとり方

号 (座標点) のとり方を示した。また 2 名の筆者による筆跡をクラスター分析し得られたデンドログラムの結果とクラスター距離 (D) を図 2 に示した。ここでクラスター距離の数値は大きなものほど筆跡のまとまりが悪い (共通性がない) ことを示している。図 2 は異者間の筆跡であるため当然クラスター距離は 201.413 と大きい。この例では両筆者間の筆跡は明確に分離されている。この方法による 30 名の筆者間の識別率は 95% であった。「愛」, 「連」, 「深」等の文字でも判別率はほぼ同様であり、クラスター分析による筆者識別はかなり有効な方法であると考えられた。

2. 因子分析

前回報告した主成分分析と同様、偏差値から計算した相関係数を利用し、筆跡の異同識別を第 1 因子パターン、第 1 因子得点により行った。今回は検査文字として「藤」を用いた。計算は主成分分析と同じ相関係数を用いて行っているため、因子分析の結果はある程度主成分分析の結果から予測できるが、軸の回転を行う因子分析にのみ見られる部分も認められた。また計算は 2 つの因子ベクトルが決定する平面内での分散が最大になるように因子軸を回転させ、いずれの平面における回転角も 0.001 ラジアン以下となった時、計算を終了させるようにした。しかし 100 回回転を行ってもこの基準が満たされないときには、共通因子はないものとして計算を打ち切った。そのため異者間の筆跡では当然であるが、同一筆者の筆跡でも記載された文字内での筆跡のバラツキが大きい場合は 100 回の回転では計算は収束せず因子分析は不可能であった。ただこのバラツキの大きさはクラスター分析から得られるクラスター距離からほぼ推定可能で



クラスター距離

5.92439 201.413  
 : MIN(17:19) \* (17:19)---(17)

図 2 クラスター分析結果  
 1-10 筆者 1  
 11-20 筆者 2  
 デンドログラム

表 1 クラスター距離と因子分析計算の可能性の関係

クラスター距離	因子分析	
	可 能	不 可 能
0- 25	4	0
26- 50	14	0
51-100	1	10
101-	1	5

ある。「藤」を用いてクラスター距離と因子分析収束の可能性との関係を検討した結果では、表 1 に示したように筆跡のバラツキが大きいことを示すクラスター距離 51 以上では二例を除いてすべて計算が収束せず分析不可能で



関数が、グループ1, 2の判別関数から大きく離れている場合は、その筆跡は両者のいずれの筆跡でもない可能性が高いということをある程度推定する必要がある。また各要素(座標点)の判別関数値の絶対値の大きな要素程両者の判別には重要な要素であることから、各要素の判別関数値からそれぞれの要素が判別にどの程度寄与しているかが明かとなる。

4. 数量化理論第II類

試料筆跡の画数が少なく要素数が12以下であるような文字では座標点を数値化しコンピュータ処理するには情報が不足し正確な結果が得られない。そこで検査者の主観が入り最上の方法とはいえないが、ある画の終筆点を はねているとか、かどげずけの仕方は鋭いか等の情報を 観察者を通じて数値化し、その結果を用いて数量化理論 第II類で筆者識別を行った。今回は試料文字に「加」を用い、同一条件で1回に10文字ずつ書かせたものを検査 試料(以下筆跡)とした。ついで検査者は個々の文字 「加」について、次の15項目のそれぞれにつき該当すると 考えた場合は1、該当しないと考えた場合は2を与える 方法で数値化した。その項目は、(1)字体は丸みを帯び ているか。(2)縦線は直線か。(3)横線は直線か。(4) 字は右上がりか。(5)字は左上がりか。(6)各画を正確 に書いているか。(7)加の「カ」の部分は丸いか。(8) 加の「カ」の部分で第1画の終筆点(1)が第2画の終 筆点(2)と比べ、垂直位置からみて上にあるか。(9) 加の「カ」の部分は丸いか。(10)加の「カ」の部分をは ねているか。(11)加の「ロ」の部分で横線が縦線より右

表3 数量化理論第II類の入力データ例

=====

<<< DATA MATRIX >>>

Table with 15 columns (1-15) and 20 rows (1-20) of binary data (0/1) representing handwriting features for the character '加'.

1~10 筆者1  
11~20 筆者2

表4 数量化理論第II類の計算結果  
1-10 筆者1  
11-20 筆者2

Two tables showing sample scores for Group 1 and Group 2. Group 1 scores range from 0.8051 to 1.1450. Group 2 scores range from -1.9385 to -0.1969. A summary table shows Group 1 has a mean of 1.0000 and SD of 0.4832, while Group 2 has a mean of -1.0000 and SD of 0.6116.

(1) 2名の筆者間の分離可能な例

Two tables showing sample scores for Group 1 and Group 2. Group 1 scores range from 0.8051 to 1.1450. Group 2 scores range from -1.9385 to -0.1969. A summary table shows Group 1 has a mean of 1.0000 and SD of 0.1061, while Group 2 has a mean of -1.0000 and SD of 0.1266.

(2) 2名の筆者間の分離不可能な例

Two sorting tables for variables 1-15. Table (1) shows a clear separation between the two groups based on variable patterns. Table (2) shows significant overlap between the groups, making them difficult to distinguish.

(1) 2名の筆者間の分離可能な例  
(2) 2名の筆者間の分離不可能な例

図4 数量化理論III類の計算結果

にてでているか。(12)加の「ロ」が縦長か。(13)加の「ロ」の大きさが「カ」の大きさ以上か。(14)加の第1画「カ」が直線か。(15)加の「カ」の部分が接しているか。の15

検出された筆跡のパターン

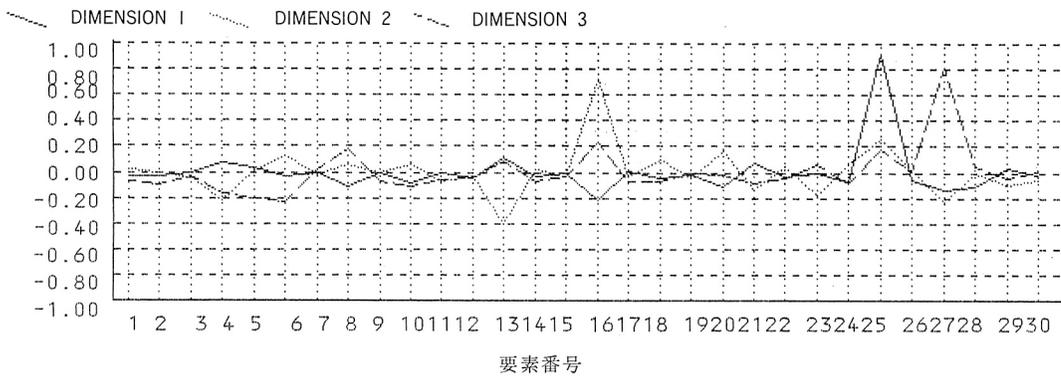


図 5 数量化理論第IV類の計算結果のパターン化

項目である。得られた検査結果の1例を表3に示した。また表3の結果を用いて数量化理論第II類で計算した結果を表4(1), (2)に示した。この結果から明かなように、この方法は判別分析である。表4(1)の例は、スコアからグループ1の6とグループ2の16は判別できない。さらに標準偏差0.4332, 0.6116から両グループは両グループ内でもかなりのバラツキがあることが明かである。またグループ2の14はスコアからグループ1に近いことを示している。次に表4(2)の例は、スコアおよび標準偏差から両グループはきわめて明確に判別できることを示している。実際に30名の筆跡について上記の方法で数値化後数量化理論第II類で計算したところ、判別率は98%であった。このことから本方法は検査者の主観が入る可能性があるものの、画数の少ない文字に対してはかなり有効な方法であると考えられた。

5. 数量化理論第III類

この方法も要素数(座標点)の少ない画数の文字について筆者識別を行う方法の一つである。方法は数量化理論第II類と同様、設定した項目について、検査者の主観により該当の有無で1あるいは2を与え、それを用いて計算した。数量化理論第II類と同様な15項目について検査文字「加」を数値化した後計算した結果を図4(1)(2)に示した。図4(1)は明確に分離した例であり、図4(2)は分離に失敗した例である。いずれの場合も sorting table で互いに近くにある筆跡ほど類似している筆跡であることから、類似性をもった筆跡間の抽出が可能である。実際の検査においては、数量化理論第II, III類の両方の計算を行い筆者識別をするのが最もよい方法と考えられた。

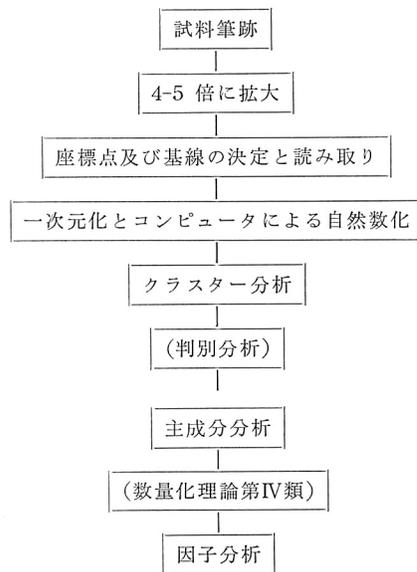
6. 数量化理論第IV類

前回報告した主成分分析に用いた相関係数を、そのまま数量化理論第IV類の計算に利用し全要素に対し関連

ない要素を抽出する方法として検討した。今回は検査文字として「藤」を用い計算したところ、図5に示したような結果が得られた。この結果から要素番号13, 25, 27が他の要素とは関連の薄いところ(「藤」を書く時それ程注意を払わないところ)であることが明かとなった。このように数量化理論第IV類は筆者の記載時の特徴を把握することができる。さらに図5のようなパターンを互いに比較することにより筆者の識別が可能となる場合が多い。

7. 筆跡の異同識別方法

実際試料の筆者識別は主成分分析, クラスタ分析, 因子分析を図6に示したような流れに従って計算した。



( )は通常は行なわない

図6 コンピュータによる筆跡の異同識別方法の流れ図

ついで計算結果を次のようにして得点として表示するようにした。クラスター分析，第1主成分得点，第1因子得点はすべて検査試料と対照試料の筆跡が分離していれば0，混合していれば1を与えることとした。ここで第1主成分得点及び第1因子得点については，検査試料と対照試料との筆跡の間の得点が最も近似した数値間で10以上離れたものを分離しているとする。

次に第1主成分パターン，第1因子パターンについて

表 5 実際試料の筆者識別例  
識別例

	クラスター	主成分パターン	因子パターン	主成分得点	因子得点
寺	混合	完全一致	完全一致	混合	混合
寺	混合	完全一致	—	混合	—
脇	混合	完全一致	—	混合	—
清	混合	不一致	一致	混合	混合
清	混合	完全一致	一致	混合	混合
博	混合	不一致	不一致	混合	分離
博	混合	一致	—	分離	—

一致度 80.7%

識別例

	クラスター	主成分パターン	因子パターン	主成分得点	因子得点
原	混合	一致	—	混合	—
原	混合	一致	—	混合	—
佐	混合	完全一致	完全一致	混合	混合
武	混合	完全一致	—	混合	—

一致度 97.1%

識別例

	クラスター	主成分パターン	因子パターン	主成分得点	因子得点
柳	混合	完全一致	完全一致	分離	分離
守	混合	一致	完全一致	分離	混合
守	混合	一致	一致	混合	混合

一致度 76.0%

識別例

	クラスター	主成分パターン	因子パターン	主成分得点	因子得点
鳥	混合	一致	完全一致	混合	混合
野	混合	完全一致	一致	混合	混合
幸	混合	完全一致	—	混合	—

一致度 96.9%

識別例

	クラスター	主成分パターン	因子パターン	主成分得点	因子得点
寺	混合	完全一致	完全一致	混合	混合
寺	混合	完全一致	—	混合	—
脇	混合	完全一致	—	混合	—
清	混合	不一致	一致	混合	混合
清	混合	完全一致	一致	混合	混合
博	混合	不一致	不一致	混合	分離
博	混合	一致	—	分離	—

一致度 80.7%

識別例

	クラスター	主成分パターン	因子パターン	主成分得点	因子得点
原	混合	一致	—	混合	—
原	混合	一致	—	混合	—
佐	混合	完全一致	完全一致	混合	混合
武	混合	完全一致	—	混合	—

一致度 97.1%

識別例

	クラスター	主成分パターン	因子パターン	主成分得点	因子得点
柳	混合	完全一致	完全一致	分離	分離
守	混合	一致	完全一致	分離	混合
守	混合	一致	一致	混合	混合

一致度 76.0%

識別例

	クラスター	主成分パターン	因子パターン	主成分得点	因子得点
鳥	混合	一致	完全一致	混合	混合
野	混合	完全一致	一致	混合	混合
幸	混合	完全一致	—	混合	—

一致度 96.9%

は，対照試料筆跡のみと対照試料に検査試料を加えた筆跡の2通りについて各々のパターンを書かせ，各要素どうしの数値を比較し，その差が絶対値で0.2以上離れている要素数の和をパターン全体の要素数の和で除す。その結果得られた数値の少数点以下第3位を四捨五入し，その数値が0.05以下ならば完全一致とし1を与える。同様に0.06以上0.15以下ならば一致とし0.8を，0.16以上0.20以下ならばほぼ一致とし0.5を，0.21以上ならば不一致として0をそれぞれ与えた。このような方法で検査筆

跡を点数として表示した。なおこれらの数値は筆者の明確な約1000の筆跡を処理した結果から経験的に算出した。得られた点数が得られるべき総点数の70%以上80%未満であれば検査試料と対照試料の筆跡はほぼ同一人の筆跡と考えられ、80%以上90%未満であれば同一人の筆跡と考えた。さらに90%以上であれば両者は同一人の筆跡と断定した。ここで因子分析が収束しないものについては、その項目はなかったものとして総点数の数から除外した。このようにして計算した数例の結果を表5に示した。この表5の文字のところに直交する直線が書かれているが、これはx軸及びy軸と基線の位置並びにその長さを示している。表5の例はいずれも長い経験に基づいた検査者が判定した結果とよく一致した。

### 結 論

筆跡の各部分を座標点として読みとり、得られた数値を用いて因子分析、判別分析、数量化理論第II, III, IV

類により識別した。さらにこれらの方法に加え、主成分分析、クラスター分析も行い、実際の検査試料筆跡についての判定方法を検討した。その結果本法による計算で筆跡の識別がほぼ満足できることが明かとなった。

なお本法による筆者識別は、長い経験と豊富な知識を必要とせず、未経験者でも経験者と同一の結果が得られた。

### 文 献

- 1) 本多正久, 島田一明: 経営のための多変量解析法, 139-153, (産業能率大学出版部), (1985).
- 2) 駒沢 勉: 数量化理論とデータ処理, 10-135, (朝倉書店), (1984).
- 3) 奥野忠一, 芳賀敏郎, 矢島敬二, 奥野千恵子, 橋本茂司, 古河陽子: 続多変量解析法, 77-217, (日科技連出版社), (1981).
- 4) 若原克文, 川村 司, 三井利幸: 応用心理学研究, 12, 13, (1987).