

# 筆跡の交点を利用した矩形診断に関する統計的分析

2008MI009 荒切彰太

指導教員：松田眞一

## 1 はじめに

これまでの研究(中村 [3], 古橋ら [6])では、個人の筆跡を分析してきた。この用途としては、重要な書類や遺書などで、本当にそれは本人が書いたものであるのかどうかを知ることが可能である。個人の筆跡は、その個人の特徴や癖があり、その個人の特徴や癖を統計的手法で量的データに変換し、個人の筆跡を判別してきた。そして、それらの研究では、矩形診断や漢字情報を用いることにより分析してきた。

文字には、交点が存在する。その文字の交点に着目して今回の方法を思いついた。本研究では、矩形診断を利用し、文字の交点を変数とすることにした。文字の交点によって、どのくらい判別率を高めることができるのかを調べたいと考え、この研究を行うこととした。

## 2 分析対象と分析方法

今回の研究では、交点を持つ5文字のひらがなを選択した。交点の数が4つ、3つ、2つ、1つとすべて含まれるようにした。「あまめみさ」、「ほもとゆね」、「みまあめさ」、「とゆねもほ」を判別する。また、「みまあめさ」と「とゆねもほ」は、「あまめみさ」と「ほもとゆね」をそれぞれ人の名前に見えないように並べたものである。

奇数行と偶数行に分割する二分分割交差確認法を使用し、線型判別関数による判別を行った。(菅 [2], Web[4],[5] 参照)

## 3 データ収集方法

5文字のひらがなを15人に10回書いてもらい、そのうち6個を分析した。集めたデータを、スキャナでコンピュータに取り込んだ。WindowsのPaintを使用して、ひとつひとつの外接長方形を作り(矩形診断)、そのドット数を長さとした。また、「あまめみさ」、「みまあめさ」、「ほもとゆね」、「とゆねもほ」のデータは、同じ15人の人物から取ったものではない。

## 4 変数説明

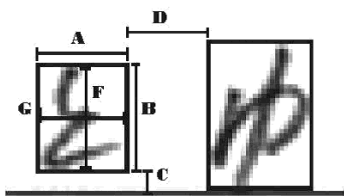


図1 変数説明

矩形診断をしたとき、漢字情報を用いた筆跡の矩形診断に関する研究(中村 [3])を参考に「縦の長さ(A)」、「横の長さ(B)」、「アンダーラインから底辺までの高さ(C)」、「隣り合う文字の間隔(D)」、「『縦』/『横』の比(E)」、「交点の縦の比率(F)」、「交点の横の比率(G)」を変数とする。

「ゆ」、「ね」は個人によって字体が異なっていた。「ほ」など、交点ができない字があった。「ほ」、「ゆ」、「ね」のそれぞれの交点の取り方は、以下のように交点を取るようになった。

## 5 交点の取り方

「ゆ」、「ね」は個人によって字体が異なっていた。「ほ」など、交点ができない字があった。「ほ」、「ゆ」、「ね」のそれぞれの交点の取り方は、以下のように交点を取るようになった。

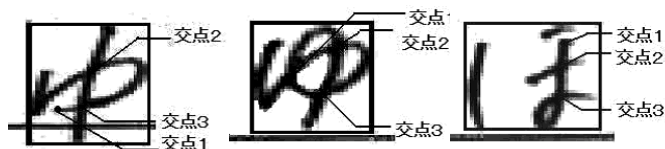


図2 ゆの交点1 図3 ゆの交点2 図4 ほの交点

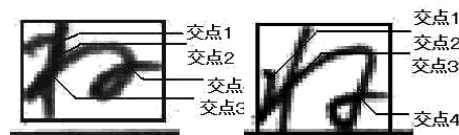


図5 ねの交点1 図6 ねの交点2

## 6 交点ありと交点なしの判別率の違い

表1 二分分割交差確認法の結果

データ	交点あり	交点なし	差
あまめみさ	91.11 %	71.11 %	20.00 %
みまあめさ	91.11 %	78.89 %	12.22 %
ほもとゆね	95.55 %	64.44 %	31.11 %
とゆねもほ	93.33 %	70.00 %	23.33 %

どの文字列も交点を変数として使用すると判別率が90%以上と高い判別率を出すことができた。交点を使用しなかったときの判別率よりも、交点を使用した方が平均して21.66%もの判別率の上昇がみられた。5文字で交点を変数として判別率を算出し、その結果がどれも90%以上の判別率を算出できるということは、個人の筆跡を5文字のひらがなで判別すると、安定した判別率を得られるということが分かった。

## 7 意味を持った文字列と意味を持たない文字列

交点ありと交点なしの判別率の結果より、意味を持った文字列と全く意味を持たない文字列の判別率の違いはないといえる。

## 8 一文字での分析

一文字で判別率を算出した場合は、70～80%の判別率が得られた。交点を4つ持つ「あ」(73.33%)よりも、交点を3つ持つ「め」(83.88%)の方が判別率が高かった。そのほかは、交点数が多いほどより高い判別率が得られた。また、一文字の判別率は、交点を使用しない場合よりも、交点を使用した方が5～22%ほど高い判別率を得られた。

## 9 交点数別の判別率の違い

交点が4つ(あ,ね)の平均の判別率は、80.278%、交点が3つ(ま,め,ほ,ゆ)の平均の判別率は、79.583%、交点が2つ(み,も)の平均の判別率は、73.332%、交点が1つ(さ,と)平均の判別率は、61.942%となった。一文字で判別したとき、交点数が多いほど判別率は高くなるということが分かった。よって、交点数が1つのひらがなを適用するよりも、交点数が4つ、3つ、2つのひらがなを適用した方が、一文字の判別の場合、より良い判別率が得られるということが分かった。

## 10 同じ交点数の判別率の違い

同じ交点数での判別率の違いは、同じ交点数であっても判別率の違いがあることが分かった。同じ交点数でも最大で20.00%ほどの判別率の差が出た。よって、一文字での判別率が高いひらがなを適用するとその文字列の判別率は上昇する傾向にある。一文字での判別率が高かったひらがなは「ほ」、「と」、「ゆ」、「ね」であった。その他のひらがなは、判別率が比較して低い、または同値であった。「あまめみさ」と「ほもとゆね」の判別率の結果と照らし合わせると、一文字での判別率が高かったひらがなが含まれる「ほもとゆね」の判別率の方が高かった。

## 11 文字数別の判別

文字数は5文字のひらがなを減らして4文字、3文字、2文字、1文字とした。そのとき、文字列が連続になるように文字数を減らした。文字数を減らして4文字、3文字、2文字、1文字の判別率は、交点ありの方が高い判別率を得られると思った。しかし、結果は3文字の判別率のみどれも交点なしの方が良い結果となった。

## 12 主成分分析を用いた判別分析の判別率

5文字のデータと3文字のデータを主成分分析でそれぞれ行ったが、何かしらの違いのある軸は見られなかった。この結果より交点ありの場合、その交点は個人の筆跡を判別にするにあたって、その個人の特徴や癖をよく反映しているのかと疑問に思った。

そこで主成分分析を用いた判別分析の判別率の算出を行う。算出方法は、主成分分析で得られた主成分得点を判別に用いるデータとし、交点ありと交点なしの次元を揃え、判別分析を行う。主成分分析で主成分得点を算出し、できるだけ少ない情報の損失でデータを縮約する主成分分析の方法は、次元を揃えるのに最適である。5文字と3文字の次元は、それぞれ10次元と6次元とした。

これらの判別率を比較し、交点が安定した変数であるのか調査する。(板野 [1] 参照)

## 13 結果と考察

表2 主成分分析を用いた判別結果

データ	交点あり	交点なし	差
あまめみさ	83.33 %	86.66 %	-3.33 %
みまあめさ	96.66 %	96.66 %	0.00 %
ほもとゆね	100 %	89.99 %	10.01 %
とゆねもほ	97.77 %	88.88 %	8.89 %
あめさ	94.44 %	93.33 %	1.11 %
とゆね	89.99 %	83.33 %	6.66 %

5文字のひらがなと「みまあめさの『あめさ』」と「ほもとゆねの『とゆね』」を主成分分析を用いた判別分析で判別率を算出した。この結果より、できるだけ少ない情報の損失で、データを縮約する主成分分析を用いたため差は大きくはないが、交点ありの方が高い判別率となった。よって、ひらがなの交点は個人の特徴と癖をよく表し、安定した変数といえることが分かった。

文字数が3文字の場合は、交点なしの場合の方が、交点ありよりも高い判別率となったが、主成分分析を用いた判別分析で判別することで、交点を変数として使用した方が判別率が高くなることが分かった。この結果も、交点が安定した変数であることを良く表している。また、3文字で判別率を算出する場合は、特殊な計算方法を行う必要があることが分かった。

## 14 おわりに

文字の交点に着目したことで、また新たな発見があった。また、データを集める際、15人に10回ずつ5文字のひらがなを書いてもらい、4セットのデータをスキャナでコンピュータに取り込み、Paintでドット数をひとつずつ測定し、18000個もの変数の数となり非常に苦労した。

## 参考文献

- [1] 板野鋭：パターン認識における主成分分析，顔画像認識を例として。  
<http://www.ism.ac.jp/editsec/toukei/pdf/49-1-023.pdf>
- [2] 菅 民郎：初心者がらくらく読める，多変量解析の実践，上，現代数学者。
- [3] 中村元樹：漢字情報を用いた筆跡の矩形診断に関する研究，南山大学数理情報学部数理科学科，2006。
- [4] フリーソフトによるデータ解析・マイニング第17回，Rと判別分析。<http://mjim.doshisha.ac.jp/R/17.pdf>
- [5] フリーソフトによるデータ解析・マイニング第24回，Rと主成分分析。<http://mjim.doshisha.ac.jp/R/24.pdf>
- [6] 古橋あい・長谷川千津・伊藤志麻・浦末直樹：統計的解析による筆跡鑑定，南山大学経営学部情報管理学科卒業論文要旨集，1996。