個人の特徴に関する矩形診断法の研究

2009SE302 :渡邉智章

指導教員：松田眞一

1　はじめに

これまでの矩形診断法を用いた研究で，個人の筆跡を分析してきた．筆跡を診断することで書類や遺書など，特定の個人が書いたものかどうかを判別すること可能である．これまで個人の筆跡に関する矩形診断法に対して，漢字情報，文字の交点を用いて研究がおこなわれてきた．(荒切 [1]，中村 [5] 参照)

人が文字を書く際には，文章で書くことが多い．文章の中には句読点があり，よく使うフレーズも存在する．また，書くものによっては丁寧に書くかどうかが変わってくる．丁寧に書いたものと判別できるのか，これらを用いてどのくらい判別率を高めることができるのかを知りたいと思い，この研究に行うことに決めた．

2　分析対象と分析方法

今回の研究では，単語ではなく文章を選んだ．人が文を書くときによく使う「私は、」と「である。」という部分に着目し，「私は、親切な人である。」と「私は、意外な人である。」の 2 つの文章を設定した．さらに，この 2 つの文章を丁寧に書いたものと普段通りに書いたものとに分けたものとに分けて，4 つのパターンの文章を分析対象とする．

これらのデータを交差確認法（leave-one-out cross-validation）を使用し，線形判別分析による判別を行う.(金 [3] 参照)

3　データの収集方法

データ収集は 2 回に分けて行った．1 回のデータ収集で，15 人の人に 1 つの文章を丁寧に 10 回書いてもらい，もう 1 つの文章を普段通りに書いてもらいデータ収集を行った．1 回目の収集から 2 週間後に 2 回目のデータ収集を行い，丁寧に書く文章と普段通りに書く文章を入れ替えて書いてもらった．集めたデータはスキャナーでコンピュータに取り込み， Windows の Paint を使用し外接長方形を作り，ドット数を長さとして数値化した．なおアンケートに協力してもらった 15 人はすべて同一人物である．

4　変数説明

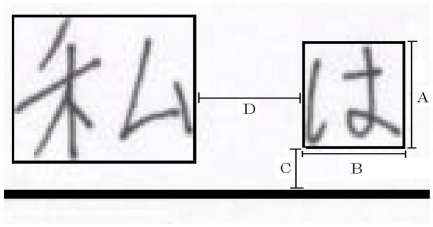

図 1　変数説明

過去の矩形診断法における研究 (荒切 [1]，中村 [5] 参照) 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)」,「隣り合う文字との間隔 (D)」の 4 変数を 1 文字の変数とする．

5　部分抽出での判別率

今回の研究で得たデータを用いて判別分析をする上で，「個人」，個人の筆跡をさらに丁寧書いたか普段通りに書いたかどうかで「個人の書き方」の群に分けて交差確認法を用いて判別を行った．用いたデータは第 1 回の筆跡，第 2 回筆跡，全ての筆跡から得た筆跡データを部分的に抽出した「私は、〜〜である。」「私は、」「である。」の部分のデータを用いた．

異なる文章から共通部分を部分抽出した場合に，誰が書いたのかを判別する個人の筆跡の判別率は「私は、」,「である。」,「私は、〜〜である。」の全ての文字列で 95% 前後と高い精度で可能である。また，誰が丁寧に書いたか普段通りに書いたかを判別する個人の書き方の判別は「私は、」,「である。」では判別率 80% 前後，「私は、〜〜である。」90% 程度の判別率を得た．矩形診断法では丁寧に書いたか普段通りに書いたかで筆跡に違いがあるが，その違いによって高い精度で判別できないことが分かった．

6　句読点ありなしでの判別率の違い

各文字列で句読点ありなしでの個人の判別の判別率を比較した．比較した結果，全ての場合で句読点ありの方が句読点なしに比べ判別率が高いことが分かった．しかし，句読点ありの場合と句読点なしの場合とでは変数の数が異なるため，適切な比較が出来るように主成分得点を用いて変数の数を 2〜8 に統一して比較した．主成分得点を用いた句点ありなしの判別率を比較し，判別率の優劣がはっきりとはみられなかった．このことから，句点は他の変数に比べ有効な変数とはなりにくい．読点ではほとんどの場合で読点ありの方が判別率が高く，読点は句点と異なり他の変数と同程度有効な変数であることが分かった．以上の結果から，句読点はひとまとまりにではなく句点・読点として全く異なる種類の文字としてみることが適切である．

7　筆跡パターン数の考察

筆跡には人それぞれ複数の筆跡パターンが存在する．丁寧書いたか普段通りに書いたかが筆跡のパターンになる考え，個人の筆跡を丁寧に書いたか普段通りに書いたかで群分けを行い判別分析を行った。その判別結果から誤判別したもののうち同じ人物が書いたものであれば正しく判別できているものとした判別率は 95% 以上と高い判別率を得られた．個人の筆跡を丁寧に書いたか普段通りに書いたか

に加え，第 1 回の筆跡か第 2 回の筆跡かで 4 群で群分けして同様に判別分析を行った．「私は、」,「である。」では共に判別率 81.1% と筆跡パターンとして捉えるには適切ではないことが分かる．

H. C. Romesburg[2] を参考に個人の筆跡をウォード法によるクラスター分析を行い 2〜8 群に群分けを行い，同様に判別分析を行った．最も高い判別率は「私は、」では群数 4 のとき 98.5%，「である。」では群数 4 のとき 95.8%，「私は、〜〜である。」では群数 3 のとき 100% の判別率となった．このことから，個人の筆跡のパターン数の最適数は 3 もしくは 4 であることが分かった．Yoshimura et. al[4] の研究の結果に近い結果となった．署名だけでなく，普通の書き文字でも筆跡パターンは 3 もしくは 4 つであることが分かった．クラスタ―分析を用いた個人の筆跡の判別は，適切な群分けを行うことで単純に個人の筆跡を判別するよりも高い精度判別できることが分かった．

8　時間経過による判別

時間経過があった場合の個人の筆跡の判別では第 1 回データを教師データ，第 2 回データをテストデータとした場合には，どの文字列でも判別率 90% 以上となった．逆に第 2 回データを教師データ，第 1 回データをテストデータとした場合には，「私は、」では 74.9%，「である。」では 81.3%，「私は、〜〜である。」では 84.6% の判別率であった．これは第 1 回と第 2 回の筆跡では違いがあったためだと考えられる．同様に個人の書き方の判別では，第 1 回データを教師データ，第 2 回データをテストデータとした場合では 55〜65% となり，第 2 回データを教師データ，第 1 回データをテストデータとした場合 49〜56% と判別率は低い結果となった．これは書くたびに人の書き方への意識の違いがあると考えられる．

9　時間による筆跡の変化

時間による筆跡の変化を調べるために各個人の第 1 回と第 2 回の筆跡の分散の変化を調べた．各個人の第 1 回での筆跡データから分散と第 2 回での筆跡の分散の変化は，「分散が大きくなった人」,「分散のあまり変わらない人」,「分散が小さくなった人」の 3 つのパターンに分かれ，その傾向は個人のよって異なる．第 1 回から時間をおいて書いた第 2 回の筆跡で筆跡の安定性は一定の法則性がないことが分かった．そこで，各文字の分散と平均の変化から筆跡の変化について考察した．平均の変化は各個人の平均の変化の絶対値の平均とした．各変数の中で最初の文字「私」の横位置が時間経過によって最も変化しやすい変数となることが分かった．また，最初の文字の横位置は最も分散の大きい変数として非常にばらつきのある不安定な変数であることが分かった．

9.1　クラスター分析を用いた筆跡の変化の考察

第 2 回データを教師データ，第 1 回データをテストデータとした場合の誤判別から筆跡の変化を考察した．判別クロス表から特定の人の筆跡に多く誤判別される人物を探し出し，クラスター分析を行い筆跡の類似性を比較した．図 2 は 3 の人の第 1 回，第 2 回の筆跡と 6 の人第 2 回の筆跡とでクラスター分析を行った結果である．

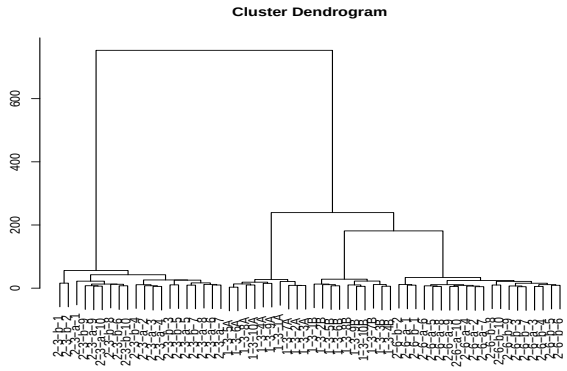

図 2　クラスター分析よる分析結果

図 2 のデントグラムから 2 群に分けた場合，3 の人の第 1 回の筆跡は 6 の人の第 2 回筆跡は同じ群となった．このように，第 1 回と第 2 回とで同じ人物の筆跡よりも他人の筆跡の方が類似性があるという結果が得られた．これは，個人差はあるが筆跡の変化によって自身の筆跡よりも他人の筆跡に非常に似通うことがあることを示す．これにより，場合によっては時間経過による筆跡の判別の判別率が下がる可能性がある．

10　おわりに

今回の研究では，丁寧に書くか普段通りに書くのかの違い・句読点の特性・時間による筆跡の変化・筆跡のパターン数などの多くの視点から筆跡について研究を行ってきた．私自身が予想していた結果とは違う結果が得られることも多く，多くの発見があった．しかし，当初予定していた漢字を部首と部外とで分割して分析する研究を時間の関係上行うことが出来なったことには残念に思う．

参考文献

[1] 荒切彰太:筆跡の交点を利用した矩形診断法に関する統計的分析，南山大学数理情報学部情報システム数理学科学科卒業論文，2012.

[2] H. C. Romesburg(西田英朗・佐藤嗣二　訳): 実例クラスター分析，内田老鶴圃，1992.

[3] 金明哲: R によるデータサイエンス，森北出版株式会社，2007.

[4] Mitsu Yoshimura，Yutaka Kato，Shinichi Matsuda，Isao Yoshimura : On-line Signature Verification Incorporating the Direction of Pen Movement，IEICE Tran，74E，2083-2092，1991.

[5] 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卒業論文，2007.