

画像情報を用いたパターンマッチングの数量化研究

- MT システムによる印鑑認証 -

日大生産工 (研) ○大坂 一司
日大生産工 矢野 耕也

概要

一般に認証や識別・同定などを目的として、画像や各種情報を用いた照合が行われている。金融機関の一部では預金者に対し、指紋認証や静脈認証などの、バイオメトリクス技術を用いた照合が行われている。ここでは品質工学における数量化技術の照合への適用を目的とし、印鑑の照合についてアプローチを行う。

1. 緒言

今日の日本文化において、印鑑は不動産取引・公正証書・自動車の購入など大きなお金にかかわる際に必要な実印や、金融機関へ口座開設時に届ける銀行印など、個人の財産に関わる重要な事象に対して大きな影響力を持っている。そのため、各種取引の際にはその印鑑が届け印と同一であるかどうかの照合が重要となってくる。そこで、本研究では印鑑の情報を数量化し、品質工学の手法による照合を行う。

2. 方法

2.1 実験手順

(1) 印鑑情報のサンプリングと数値化

印鑑を用紙に押印し、スキャナでデジタル画像へ変換した。その後、デジタル画像のRGB カラーモデル数値情報を収集した。その情報のうち、最も感度の高かった青の数値情報を使用した。

(2) 数値情報の整理

印鑑の形の情報をより明確にするため、収集した数値のうち、200 以下を 1、200 以上を 0 とした。また、サンプリング時に生じたズレを補正するためデータの開始点を統一し、文字の角度を調整する処理を行った。

(3) 特徴の抽出

解析に使用した特徴は図 1 のように、各行及び列の 1 の和と、連続している 1 のグループ数を特徴として抽出した。

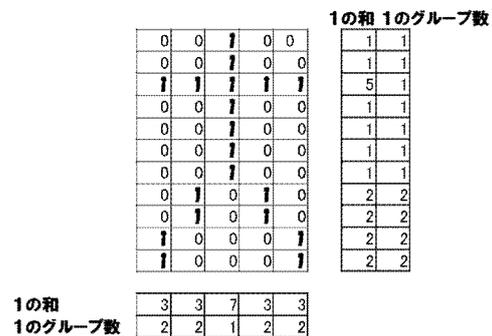


図 1 特徴の抽出

(4) RT 法による解析

解析方法として、MT システムの中でパターン分析に適している RT 法を用いて解析した。

2.2 RT 法とは¹⁾

RT 法では標準 SN 比 η と比例定数 β に情報を圧縮し、それぞれを項目として用いて二項目から求めた距離の値により、パターン認識を行う

Research of making to amount on pattern match using image information

-Stamp attestation by MT system-

Hitoshi OSAKA, Koya YANO

手法である。その際、目的に対してパターンが均質な基準となる集団を単位空間として定義し、対象データと単位空間の中心からの距離を求めるものである。また、情報を β と η に圧縮するため、大量のデータを扱うことができ、単位空間のメンバー数が特徴の数より少なくても計算可能という利点がある。

2.3 RGB カラーモデル²⁾

RGB カラーモデルにおける色は、赤・緑・青の各要素がどれだけ含まれているかで記述することができる。コンピュータにおける色の表示には、各要素の明度を 0 から 255 の数字で表現しており、本研究ではこの数値を利用して解析を行った。

2.4 画像サンプリング

本研究では、図 2 に示した 3 種類の印鑑を使用した。まず、基準として「大坂」の印鑑 A、判別の対象として印鑑 A とは別の「大坂」の印鑑 B と、印鑑 B と同じ書体で類似文字「大城」の印鑑 C を用意し、印鑑 A で 200 枚、印鑑 B・印鑑 C で 20 枚ずつサンプリングした。



図 2 使用した印鑑

2.5 文字の角度補正²⁾

文字の角度補正には、画像回転処理の基本アルゴリズムを使用した。入力点の座標を (x_1, y_1) 、中心を (cx, cy) 、角度 θ 回転させる場合の出力点の座標を (x_2, y_2) 式(1)及び式(2)で算出することができる。

$$\begin{aligned} x_2 &= (x_1 - cy) * \cos(-\theta) \\ &\quad - (y_1 - cy) * \sin(-\theta) + cx \quad (1) \end{aligned}$$

$$\begin{aligned} y_2 &= (x_1 - cx) * \sin(-\theta) \\ &\quad + (y_1 - cy) * \cos(-\theta) + cy \quad (2) \end{aligned}$$

しかし、この計算方法を使用すると、入力画像の位置から出力画像内の座標を計算してい

るため、穴が生じてしまう。そこで、図 3 のような出力画像の座標に対する入力画像の座標を逆算する。

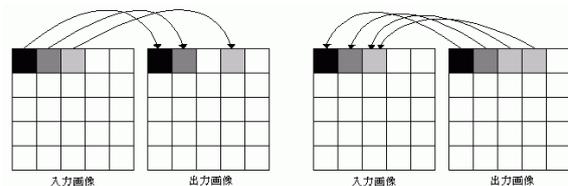


図 3 画像回転処理における入出力対応

このとき用いる計算式は式(3)及び式(4)である。

$$\begin{aligned} x_1 &= (x_2 - cy) * \cos(\theta) \\ &\quad - (y_2 - cy) * \sin(\theta) + cx \quad (3) \end{aligned}$$

$$\begin{aligned} y_1 &= (x_2 - cx) * \sin(\theta) \\ &\quad + (y_2 - cy) * \cos(\theta) + cy \quad (4) \end{aligned}$$

角度 θ の算出方法においては、印鑑情報の中心の平均値を中心 (cx, cy) 、図 4 のポイントを基準点 (x_3, y_3) とし、式(5)で求めた。

$$\theta = \tan^{-1} \left(\frac{y_3 - cy}{x_3 - cx} \right) \quad (5)$$

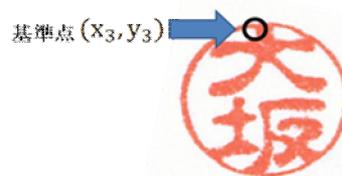


図 4 基準点のポイント

3. 結果

3.1 印鑑 A を単位空間とした場合(結果 1)

印鑑 A のデータ 200 枚を単位空間、基準となる印鑑 A で 200 枚、印鑑 B・印鑑 C で各 20 枚のデータを対象データとし、解析した結果を図 5 に示した。図 5 を見ると、印鑑 A とそれ以外の印鑑では距離に大きな差が生じ、同じ文字の印鑑でも印鑑の文字の書体が違えば判別することが可能であることがわかった。しかし、印鑑 B と印鑑 C では判別が困難であった。

3.2 単位空間の再選定(結果 2)

より識別精度を上げるために、単位空間データの再選定を行った。単位空間の選定方法として、基準とした印鑑 A のデータについて式(6)

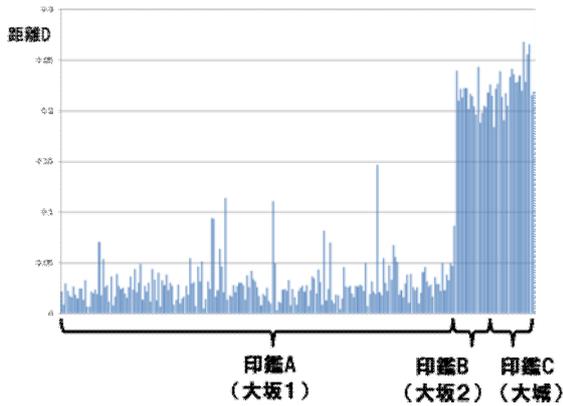


図5 印鑑Aを単位空間とした場合の距離D

より分散 v_0 を求める。次いで印鑑A 200枚のデータそれぞれについての誤差分散 v_e を式(7)～式(10)より求め³⁾、誤差分散の比 v_e/v_0 を算出し、値が1に近いもの、すなわち、項目データのばらつきが分散 v_0 に近くて小さいものを10個選び、改めて単位空間とした。

$$V_0 = \frac{(x_{11}-m_1)^2 + \dots + (x_{nk}-m_k)^2}{nk} \quad (6)$$

$$S_T = (y_{11} - m_1)^2 + \dots + (y_{nk} - m_k)^2 \quad (7)$$

$$S_m = \frac{[(y_{11}-m_1) + (y_{12}-m_2) + \dots + (y_{1k}-m_k)]^2}{k} \quad (8)$$

$$S_e = S_T - S_m \quad (9)$$

$$V_e = \frac{S_e}{k-1} \quad (10)$$

結果的に、分散を小さくした基準パターンを用いて再解析を行ったことになり、その結果を図6に示した。図6を見ると印鑑Bと印鑑Cにも差が生じた。今回用意した印鑑は、印鑑Aと印鑑Bが同じ文字、印鑑Bと印鑑Cが同じ書体であることから、理想的な解析結果は、距離Dが印鑑A<印鑑B<印鑑Cとなることである。単位空間を誤算分散の比の大ききで選定し直したことにより、この条件を満たす結果を得ることができた。

3.3 単位空間の絞り込み(結果3)

3.2では誤差分散の比 v_e/v_0 で単位空間を選定することにより、識別精度を向上させることが可能であったことを示したが、銀行の口座開

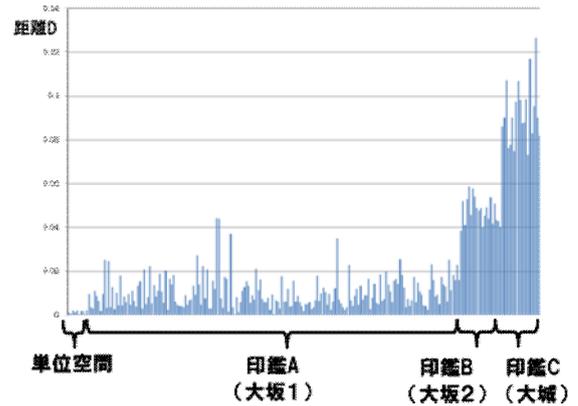


図6 単位空間を選定した場合の距離D

設などでは、届け出印の登録の際には一度の押印で、それを基準とする場合が多い。そこで、ごく少数のサンプルによる単位空間では、どの程度の識別精度を得られるか否かを検証をする。まず、多変量解析の特性として、単位空間のデータ数が一つでは解析を行うことができない。しかし、RT法では各項目の平均値から求めた β と η を項目とするため、単位空間データが二つあれば解析が可能である。そこで、印鑑Aで新たに二つのサンプリングをした。それらを新たに単位空間として解析を行い、その結果を図7に示す。

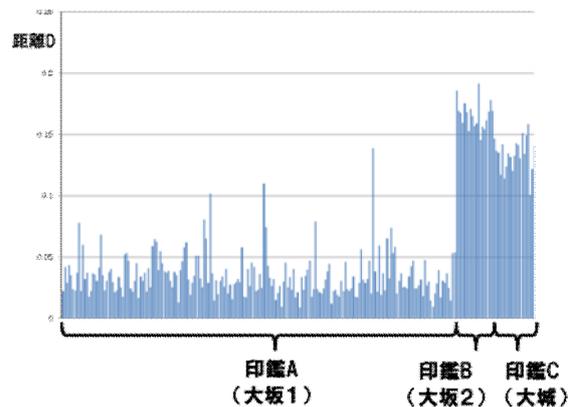


図7 単位空間データが二つの場合の距離

図7では、印鑑Bと印鑑Cの距離Dの関係が結果2とは逆転したが、印鑑Aとそれ以外の印鑑とでは差が表れ、少ない基準データでも正しく判別を行うことができた。

3.4 識別力の比較

単位空間の選定方法による識別力を比較す

るため、各結果について、単位空間、印鑑 A、印鑑 B、印鑑 C の平均距離と、印鑑 A を基準とした距離の比を表 1 にまとめた。

表 1 各結果の平均距離と印鑑 A との比

	平均距離				印鑑Aとの比		
	結果1	結果2	結果3		結果1	結果2	結果3
単位空間		0.0013		単位空間		0.126	
印鑑A	0.0290	0.0105	0.0348	印鑑A	1.000	1.000	1.000
印鑑B	0.2124	0.0480	0.1652	印鑑B	7.317	4.577	4.750
印鑑C	0.2295	0.0924	0.1331	印鑑C	7.905	8.808	3.826

結果 1 と結果 2 を比較してみると、結果 2 では印鑑 B と印鑑 C とでは印鑑 A との比に差が出ているが、印鑑 B の比は結果 1 では 7.3 倍に対して、結果 2 は 4.6 倍と識別精度が下がったが、印鑑 C については 7.9 倍から 8.8 倍に向上している。

結果 3 においては、いずれの印鑑の識別はできているものの、結果 1 に比べて印鑑 B、印鑑 C 共に識別精度が下がっている。また、印鑑 B と印鑑 C の識別精度は、結果 1、結果 2 のものとは逆転し、印鑑 C の方が印鑑 A に近い距離となった。つまり、印鑑 C のパターンは、印鑑 B より印鑑 A のパターンに数値上は類似しているものと考えられることができる。

4. 考察

単位空間データの選定方法により、印鑑 B と印鑑 C の距離 D の関係性に大きな違いが生じた。単位空間を厳密に選定した結果 2 では理想的な結果を得ることができたことから、結果 1 では単位空間のデータ数が多すぎ、ばらつきの許容幅が広がったのが結果 1 の識別性の低さの原因であることが考えられる。また、結果 3 では、印鑑 B と印鑑 C の理想的な距離 D の関係とは逆の結果となった。原因として、単位空間が少なすぎるため印鑑 A の特徴を示す項目の情報不安定であったこと、または、単位空間のサンプルが均質な基準ではなかったことなど考えられる。このように、0・1 のような二値データは平均値が 0.5 となり、0 や 1 の値に近くならないので少数のデータでは不安定な

結果を与えることがある。また、それぞれの結果において、印鑑 A でも大きく距離が出ているものがいくつか存在した。大きく距離が出ているサンプルすべてに、図 8 のような滲み、欠け、ゴミ付加などが見てとれたため、これらが距離 D の値を大きくする誤差の原因であると考えられる。

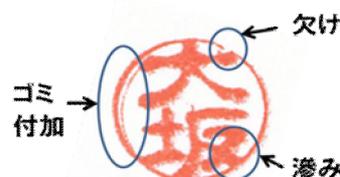


図 8 誤差の要因

5. まとめ

本研究では、RT 法により印鑑の照合ができる可能性を示すことができた。

一般に印鑑の照合においては、基本的に単位空間サンプルが一つしかない場合が多い。本研究では解析方法の性質により、最低二つの単位空間サンプルにより解析を行ったが、これを一つの単位空間サンプルで行い、識別精度を上げていく技術が必要であると考えられる。

RT 法による印鑑認証をシステム化するためには、多種の印鑑で照合が可能か否かの検証と、データ量の圧縮、また今回手作業で行った基準点のポイントを自動で行うための技術などが必要である。また、印鑑だけでなく、指紋の認証、筆跡の鑑定、電子回路の不良品判別やモノの形の照合など、幅広い応用が可能であることが考えられる。

「参考文献」

- 1) 手島 昌一、入門 MT システム、日技連出版社(2008)、pp79～90
- 2) デジタル画像処理編集委員会、デジタル画像処理、CG-ARTS 協会、(2009)、pp29、pp157
- 3) 矢野 宏、品質工学計算法入門、日本規格協会、(2000)pp21～27