

人間乱数によるSOMを利用した個人識別の可能性

田中 侑希^{1,a)} 田中 美栄子^{1,b)}

概要: 人が生成した乱数列のことを人間乱数と呼び、被験者の体調や生成方法によって数列の特徴が変化することが知られている。人間乱数を個人で識別するためには乱数列に個人の癖(特徴)が反映される生成方法を考える必要がある。本研究では乱数列に個人の癖が反映される生成方法として、乱数生成時の思考度合が異なる2種類のデータを採取し比較を行うことで両者の特徴を捉えることを目的とした。採取したデータを用いて様々な指標を計算した結果、直感的に生成した場合と熟考して生成した場合とで明確に差が現れず、どちらもある程度個人の癖が反映されることがわかった。そこで8種類の特徴量となる指標を用いてSOMによる個人分類実験を行ったところ、熟考データに比べて直感データの方が個人を分類する上で、より適していると思われる結果となった。また、特徴量のうち、データ採取時の被験者の行動パターンを反映するものが重要であったことから、データ列の解析だけでなく被験者の行動を指標として取り入れることによって個人分類の可能性が高まることがわかった。

キーワード: 人間乱数, SOM, 個人識別

Possibility of the Personal Identification Using SOM by a Human Random Generation

YUUKI TANAKA^{1,a)} MIEKO TANAKA-YAMAWAKI^{1,b)}

Abstract: The human random numbers are the random sequences generated by human, which are known to reflect the subjects condition and also to depend on the methods of data takings. In order to identify individuals based on such data, we need a good experimental method in which the feature of individual subjects is reflected in the data as much as possible. We aimed in this work to compare two methods of data taking, the 'intuitive' generation and the 'deliberate' generation. We found that the difference between the results of the two types of data generation is rather small, and both reflect the features of individual to some extent. The result of SOM learning based on eight indicators, however, shows that the 'intuitive' data are more suitable for individual identification than the 'deliberate' data. We also found that the two indicators out of the eight that are related to the subjects habitual pattern play an important role in the SOM classification, indicating the importance of taking account of subjects habitual pattern as well as the features exhibited in data sequences.

Keywords: Human Random Generation, Self-organized maps, Personal identification

1. はじめに

近年、コンピュータや携帯電話の普及により身近に個人認証システムを利用する場面が増えてきておりその必要

性が増している。本稿では個人認証に人間乱数を利用することを考えたい。個人認証とは「本人しか持ち得ない属性を元にその属性を確認し本人であることを証明すること」である。現在では本人しか持ち得ない属性として知識属性(パスワード方式)、所有物属性(クレジットカード、ICカード等)、生体属性(指紋、網膜、声紋等)が利用されている。それぞれの属性を使用する際の問題点として知識の

¹ 鳥取大学大学院工学研究科エレクトロニクス専攻
Tottori University, Graduate School of Engineering, Department of Information and Electronics

a) s072032@ike.tottori-u.ac.jp

b) mieko@ike.tottori-u.ac.jp

忘却, 所有物の紛失, 生体情報登録への抵抗や登録情報の変更不可等が挙げられる。そこでこれらの問題点を補う方法として人間乱数を利用することを考えた。人間乱数を利用できれば暗記や物の所持が必要なく, 登録への抵抗もないため大変有効な個人認証システムとなる。

人に「デタラメになるように数字を並べてください」と言う作業をしてもらう。この作業のことを人間乱数テストや乱数テスト, 乱数発生テスト等と呼び, 作成された数列を人間乱数(列)と呼ぶ。得られた数列は完全な乱数列ではなく生成者の癖や状態が反映された特徴的な数列になることが知られている [1]。また, 同じ人が生成した数列であっても生成方法によって乱数の性質が変化することも分かっている [2][3][4][5][6]。そこで人間乱数を個人認証に利用するために, 個人の癖(特徴)が数列に反映される生成方法を探ることとした。人の癖は直感的な行動を取ったときに現れると考えた。なので, 乱数を直感的に考えることで数列に癖が反映されると考えた。本研究は, 直感的に考えて生成した乱数と熟考して生成した乱数の思考度合が異なる2種類の人間乱数を様々な指標を用いて比較し数列の特徴を捉える。また SOM を用いてデータの個人分類が可能であるか検証することで人間乱数による個人識別の可能性を示すことを目的とする。

2. 乱数生成方法

個人認証システムへの応用という目的から数列に個人の癖が反映される方法を探るため, 乱数を直感的に生成する方法と熟考して生成する方法を提案する。提案採取手法について条件をまとめたものが表1である。直感と熟考に共通する条件として, 使用するシンボルは一般的に人間乱数テストで用いられる0~9の整数, 乱数列の長さは50, データの採取にはコンピュータにテンキーを接続したものを記録装置として使用する。乱数列の生成は直前に入力した数が不可視となるプログラムを使用して被験者自身が入力を行う。直前の入力を不可視にした理由はデータに脳の状態を反映させるために, 視覚情報ではなく脳内の記憶を利用してもらうためである。プログラムは最初にユーザ名を入力し, その後被験者の任意のタイミングで入力を開始し, 乱数を50個入力した所で自動的に決められた書式のファイルを出力するものを作成した。また, 制約条件として直感と熟考の思考度合の異なる生成条件をそれぞれ設けた。直感的と判断する目安として入力間隔を2秒以内とした。これは五味等の報告 [3] より, 生成速度が5秒間隔以上と1秒間隔以下で生成したデータに大きな違いが見られた。このことから, 1秒間隔以下を直感的に生成したデータと判断する目安にした。しかし1秒間隔以下の入力は時間を意識してしまい焦りから正しく被験者のデータが生成されないと考え, 時間を意識しなくても生成できる入力間隔として2秒間隔以内とした。但し思考度合が重要であるため時

表 1 提案採取方法の条件一覧

Table 1 The data taking conditions

| 条件 | 値 | |
|--------|-----------------------------|-------------------------|
| | 直感 | 熟考 |
| テスト説明 | 深く考えず直感的に乱数になるように数を入力してください | よく考えて乱数になるように数を入力してください |
| 制約条件 | 直感的に判断して乱数生成(目安として2秒以内の間隔) | 熟考して乱数生成 |
| シンボル | 0~9の整数(10種類) | |
| 乱数列の長さ | 50 | |
| 生成方法 | 直前の数が不可視となるプログラムを使用 | |
| 記録方法 | テンキーを接続したPC | |
| 記録者 | 被験者 | |

間に着いては気にしないよう被験者に強く説明した。熟考の場合については思考時間が人によってバラバラであるため特に制限を設けていない。また, テスト開始前の被験者への説明は簡単にプログラムの使い方を説明した後にそれぞれ「深く考えず直感的に乱数になるように数を入力してください」, 「よく考えて乱数になるように数を入力してください」と指示した。乱数がどのようなものかについては個人で考えてもらうようにした。これは乱数がどのようなものを説明するのが困難であり, 説明の仕方によって被験者に誤った認識を与えてしまうことを防ぐためである。こうして得られたデータは1試行(長さ50のデータを1個生成)を1ファイルとして扱う。

2.1 本研究で使用するデータ

本研究で使用したデータは, 同大学の工学部情報系学科に所属する21~23歳の男性6名(A~F)から直感で採取したデータを一人当たり50ファイル, 熟考で採取したデータを一人当たり30ファイル採取した。またデータ採取は一度に最大10試行までとし, 採取日時は各々の判断で比較的気持ちが落ち着いているときに被験者のみで行った。ここで直感的に考えて生成した乱数列を直感人間乱数, 熟考して生成した乱数列を熟考人間乱数と呼び以降区別する。また乱数の比較用に擬似乱数としてrand()等で用いられる線形合同法(LCG)のデータ(Seed:1~50)も用意した。

3. 乱数列の特徴を捉える指標

乱数列から特徴を捉える際ただ数列を眺めてもよくわからない。そもそも乱数とは本来規則性のないものである。そこで乱数列を比較するために数列の規則性を数値化できる指標を用いて乱数列の特徴を捉える。今回解析に使用した指標は8種類である [6][7][8]。各指標については3.1~3.8節で説明する。また, 各被験者の人間乱数に機械乱数を加えた7者に対して各指標の平均値をまとめたものを表2表

表 2 直感人間乱数の被験者 6 名 (A~F) と LCG (L) の各指標 (H, ..., TIME) の各 50 データの平均値/標準偏差

Table 2 The average / standard deviation of the values of 8 indices (H, ..., TIME) over 50 files of 'intuition' data are shown for 6 subjects (A~F) and LCG (L)

| | H | ADJ | TPI | PL | RG | RP | TKD | TIME |
|---|---------------|--------------|----------------|---------------|---------------|---------------|--------------|-------------|
| A | .978 /.017 | 7.4 /4.5 | 114.3 /8.7 | .16 /.54 | 9.2 /.58 | 47.4 /10.4 | 56.0 /2.5 | .54 /.07 |
| B | .991 /.004 | 14.5 /5.1 | 106.3 /10.9 | .36 /.86 | 9.77 /1.17 | 44.6 /8.2 | 51.9 /4.0 | .97 /.22 |
| C | .983 /.010 | 24.9 /6.8 | 86.8 /10.5 | 1.97 /1.72 | 9.39 /1.30 | 40.4 /6.9 | 46.1 /3.7 | .81 /.20 |
| D | .970 /.016 | 23.1 /6.2 | 89.9 /10.6 | 1.18 /.86 | 8.89 /1.38 | 48.2 /8.2 | 50.3 /3.8 | .33 /.06 |
| E | .949 /.028 | 26.1 /8.3 | 87.3 /12.8 | 1.38 /1.98 | 8.77 /1.55 | 52.3 /9.7 | 50.9 /6.1 | .44 /.13 |
| F | .978 /.011 | 26.8 /6.4 | 94.6 /9.7 | 1.46 /1.62 | 9.27 /1.42 | 45.7 /9.5 | 44.9 /3.1 | .49 /.04 |
| L | .959 /.019 | 17.2 /5.1 | 96.5 /8.5 | .83 /1.25 | 7.79 /1.65 | 36.5 /6.8 | 48.2 /3.7 | |

表 3 熟考人間乱数の被験者 6 名 (A~F) の各指標 (H, ..., TIME) の各 30 データの平均値/標準偏差

Table 3 The average / standard deviation of the values of 8 indices (H, ..., TIME) over 30 files of 'deliberate' data are shown for 6 subjects (A~F)

| | H | ADJ | TPI | PL | RG | RP | TKD | TIME |
|---|---------------|--------------|---------------|---------------|---------------|---------------|--------------|--------------|
| A | .991 /.006 | 4.6 /2.8 | 119.5 /7.5 | .07 /1.35 | 9.72 /1.28 | 46.0 /10.2 | 59.2 /3.0 | 1.57 /.45 |
| B | .990 /.007 | 19.6 /6.1 | 104.5 /7.7 | .72 /1.08 | 9.45 /1.34 | 46.4 /8.6 | 52.7 /3.2 | 1.20 /.37 |
| C | .986 /.010 | 17.7 /5.3 | 95.9 /8.8 | 1.05 /1.51 | 9.38 /1.38 | 40.4 /8.6 | 56.7 /3.2 | 1.34 /.41 |
| D | .978 /.014 | 15.2 /4.9 | 104.7 /8.2 | .46 /1.84 | 8.94 /1.47 | 39.3 /8.9 | 50.9 /3.7 | 1.61 /.82 |
| E | .944 /.025 | 14.7 /4.9 | 105.1 /9.5 | .33 /0.74 | 8.33 /1.53 | 52.9 /8.5 | 60.0 /5.7 | 1.82 /.72 |
| F | .985 /.006 | 23.7 /5.7 | 93.0 /10.4 | 1.12 /1.32 | 9.46 /1.28 | 43.5 /6.9 | 49.1 /2.5 | 0.70 /.12 |

3 に示す。表 2 は直感人間乱数 50 ファイルの平均値を示し、表 3 は熟考人間乱数 30 ファイルの平均値を示す。また各平均値の次の段に標準偏差を示した。

3.1 Entropy (H)

Entropy はシンボルの生成偏差を表し、次式で定義される。(シャノンエントロピー)

$$H = - \sum_i P_i \log P_i \quad (1)$$

この時、 P_i は i 番目のパターンの出現確率を示す。ここでは $i=0\sim 9$ の値を取り、一桁の数の生成偏差を求めている。H は $0\sim 1$ の値を取り 1 に近い時シンボルの生成偏差が均

一であることを示し、逆に 0 に近いほど偏った生成偏差を示す。一様乱数である場合、有限の区間で区切られた区間内の実数は同じ確率で現れる。つまり 1 に近いほど乱数度が高いと言える。

3.2 Adjacency (ADJ)

Adjacency とは隣接する文字の演算差の絶対値 1 が出現する確率を表すもので次式で定義する。

$$ADJ = 100 \times \frac{NAP}{m-1} \quad (2)$$

m は数列の長さを表しここでは $m=50$ である。NAP は隣接する数字の絶対差 1 の個数を示す。ADJ は $0\sim 100$ の値を取り 0 に近いほど出現確率が低いことを示し、100 に近いほど出現確率が高いことを示す。

3.3 Turning Point Index (TPI)

Turning Point Index とは数列の上昇と下降が切り替わるポイント Turning Point (TP) の出現回数を式 3 の期待値と比較した値で定義される。

$$TP_{expected} = \frac{2}{3}(m-2) \quad (3)$$

$$TPI = 100 \times \frac{TP_{observed}}{TP_{expected}} \quad (4)$$

$TP_{observed}$ は実際に測定した TP の出現回数、 $TP_{expected}$ は乱数の TP 出現回数の期待値を示す。すなわち TP の出現回数が多いと 100 以上の値になり逆に少ないと 100 以下の値となる。ここで例として数列「5,3,4,6,2,8,9,7,1,0」があったとすると TP は「3」「6」「2」「9」と 4 回出現したので $TP_{observed} = 4$ となる。

3.4 Phase Length (PL)

Phase Length とは TP が発生する間隔距離 d の出現回数の期待値に対する比で表す。

$$PL(d)_{expected} = \frac{2(m-d-2)(d^2+3d+1)}{(d+3)!} \quad (5)$$

$$PL(d) = \frac{PL(d)_{observed}}{PL(d)_{expected}} \quad (6)$$

$PL(d)_{observed}$ は実際に測定した $PL(d)$ の出現回数、 $PL(d)_{expected}$ は $PL(d)$ の期待値を示す。この時、 m は数列の長さを表しここでは $m=50$ である。 d は TP が発生する間隔の長さで $1\sim m-3$ である。ここで例として数列「2,3,5,4,5,6,7,8,6,1,3」があったとすると TP は「5」「4」「8」「1」である。この時各 TP 間の長さは「1」「4」「2」となる。よってこの数列の $PL(d=1,2,4)_{observed} = 1$ となり、それ以外の $PL(d)_{observed}$ は 0 となる。

$m=50$ の時の各 d の期待値を求めたものが表 4 である。

表 4 PL(d) の期待値 (m=50)

Table 4 The expected value of PL(d) (m=50)

| d | 期待値 | d | 期待値 |
|---|-----------|-------|----------|
| 1 | 19.583333 | 6 | 0.012731 |
| 2 | 8.433300 | 7 | 0.001604 |
| 3 | 2.375000 | 8 | 0.000178 |
| 4 | 0.506349 | 9 | 0.000018 |
| 5 | 0.087450 | 10 47 | 0.000000 |

表 5 RD の出現回数

Table 5 The number of occurrences of the RD

| 繰返し距離 d | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---------|---|---|---|---|---|---|---|---|
| 出現回数 | 1 | 1 | 1 | 0 | 0 | 2 | 0 | 0 |

これをみると d=5 から期待値が大幅に減少する。これは数列の長さによる影響で m に比例して d の期待値も増減する。今回 d=5~47 は使用しないこととした。また d=1,2,3 についてもあまり有効な特徴が見られなかったためここでは d=4 を使用する。

3.5 Repetition Gap (RG)

Repetition Gap とは同じシンボルが繰り返し出現する距離 Repetition Distance(RD) の値を解析する指標である。今回は RD の平均値 (mean) を使用する。ここで例として数列「2,3,7,8,8,7,2,3,2」があったとするとシンボル「2」は 6 個先とさらにその 2 個先に出現する。シンボル「3」は 6 個先にシンボル「7」は 3 個先、シンボル「8」は 1 個先に出現する。これをまとめると表 5 のようになる。この表から繰返し距離の平均値を求めた値が RG である。

3.6 Repeat Pattern (RP)

Repeat Pattern とはデータ中における隣接する 2 文字の繰返し出現する頻度を表したもので次式で定義される。

$$RP = 100 \times \left(1 - \frac{NRS}{m-1}\right) \quad (7)$$

m は数列の長さを表しここでは m=50 である。NRS は一度しか出現しなかったパターン個数を示す。RP は 0~100 の値を取り、0 に近いほど繰返しのパターンが少なく、100 に近いほど繰返しのパターンの生成が多いことを示す。

3.7 Ten Key Distance (TKD)

Ten Key Distance とは数を入力する際のキーの総移動距離を表したものである。今回はプログラムの仕様上数を打つ度にエンターキーを押さなければならなかったため、実際に指が移動した距離ではなく入力したキーの配置を元に計算した値である。キー間の距離を図 1 のような座標に置き換え入力したキーとその前に押されたキーとの距離を計算し合計したものである。0 については 2 つ分のスペース

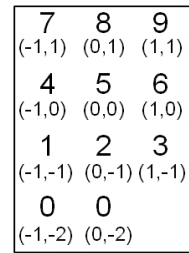


図 1 テンキーの座標

Fig. 1 Coordinate on the ten key

があるため前後のキーとの最短距離になるように定めた。例えば数列「2,4,5,9,1,7,0,6」があったとき、それぞれの移動距離は $\sqrt{2}, 1, \sqrt{2}, 2\sqrt{2}, 2, 3, \sqrt{3}$ となる。

3.8 生成時間 (TIME)

乱数を生成する際に生成間隔も同時に記録した。このデータを使い生成間隔の平均値を求めた。

4. 個人の癖の存在

人間乱数に個人の癖がどのように反映されているか検証する。表 2 を見ると、H は若干であるが熟考に比べて直感の方が低い値を取る傾向がある。これは熟考することによって僅かであるが複雑な乱数を生成していることが分かる。ADJ は熟考に比べて直感の方が大きな値を取る傾向がある。これは人が 0-1 や 7-6 など連続した数字を避ける傾向があるため、熟考した場合は低くなったと考えられる。TPI は熟考に比べて直感の方が低い値を取る傾向がある。これより熟考することで TP を意識的に増やそうとしているのが分かる。PL は熟考の場合個人間であり大きな差は見られないが直感の場合大きな差が見られた。RG は直感と熟考でどちらも同じような値を取り、また個人間での差もあまり見られない。RP も直感と熟考であり差は見られない。TKD は熟考に比べて直感の方が低い値を取る傾向がある。これは熟考の場合は近くにあるキーを押すのを避ける傾向があるのに対し直感的にキーを押すことでキー配置よりも乱数生成に意識が行き結果として熟考より低い値になったと考えられる。また直感 LCG の値に近いことからキー配置よりも乱数生成に意識が働いていたと言える。TIME は直感ではほぼ 0.5~1 秒以内であるのが分かる。熟考は平均すると 1.5 秒程度であるのが分かる。

以上のことより直感人間乱数と熟考人間乱数とであまり大きな変化は見られないが、ADJ, TPI, PL, TKD など思考度合による差は見られた。また、思考度合によって値は変化するが直感と熟考で個人間の差は同じ傾向が見られた。しかし、誤差を見ると熟考に比べ直感の方が全体的に大きくなっている。これは熟考の場合安定して複雑な乱数を生成するのにに対し直感生成時期の違いによって性質の異なる乱数を生成していると考えられる。

表 6 図 4 における集合率

Table 6 The aggregation rate in a Fig. 4

| 被験者 | A | B | C | D | E | F | Ave |
|---------|----|----|----|----|----|----|-----|
| 集合率 (%) | 98 | 90 | 90 | 90 | 75 | 85 | 88 |

5. SOM を利用した人間乱数の分類

本研究で扱う 8 次元データの分類方法として、全指標が互いに独立とは限らない場合にも応用でき、また結果を分かりやすく可視化できる利点を持つ SOM を利用する。

5.1 SOM とは

自己組織化マップ (SOM : Self-Organizing Maps) はニューラルネットワークの一種で、多変量の特徴をもつデータ群を性質の類似度によって分類する教師なしクラスタリングである [9][10][11][12]。

5.2 直感人間乱数と熟考人間乱数の個人分類実験

直感人間乱数と熟考人間乱数がそれぞれ個人で分類できるか実験を行った。直感人間乱数を一人各 40 ファイル使用して学習したマップを図 2 に、熟考人間乱数を一人各 30 ファイル使用して学習したマップを図 3 に示す。また、それぞれの学習パラメータは量子化誤差が小さくなるよう実験的に求めたものを使用した。

図 2 を見ると個人毎にデータが集まっているのが分かる。しかし、よくマップを見るとデータ間の距離が全体的に近いことが分かる。特に B と C, C と F, D と E のように近くにあるデータは一部が混ざり合うように分類されている。次に図 3 を見ると A, E は個人のデータが集まっているのが分かる。しかし、B, C, D, F は一部集まって分類されているが多くは混ざりあって分類されている。よって熟考人間乱数での個人識別は難しく直感人間乱数は個人識別の可能性はある。

ここで、直感人間乱数がどの程度個人分類できているかを分かりやすくするため図 2 の学習マップ内に境界線を引き個人の領域に分割した。境界線は個人の集合が最大になるように、かつ最短でシンプルになるように引いた。境界線を引いたマップを図 4 に示す。図 4 を見ると個人でデータが集まっているのがより分かる。また、各境界線内の領域を各個人の領域 (図 5) とし、各 40 ファイルが同一人物の領域に分類された割合を集合率として結果を表 6 にまとめた。表 6 を見ると A は 98% とほぼ全てのデータが個人で分類されているのが分かる。B, C, D, F についても 9 割程度集合しておりデータが個人毎に分類されているのが分かる。E については 75% と少しデータが散らばっているが個人で概ね分類できている。

次に図 2 上のマップが正しく学習されているかを調べるため個人識別実験を行った。使用したデータは学習で使用



図 2 直感人間乱数 240 ファイルの学習マップ

Fig. 2 SOM map of the 240 'intuition' data

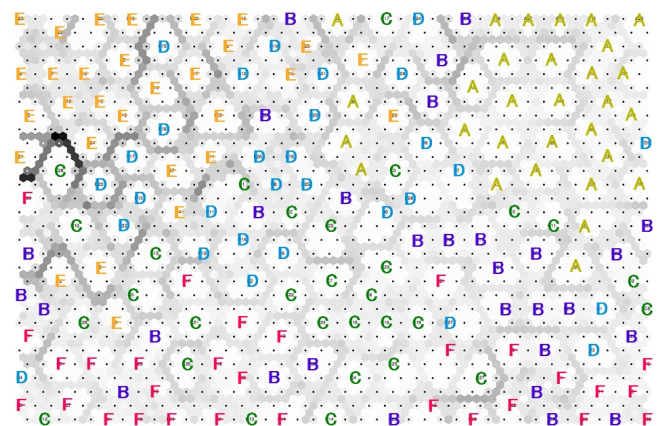


図 3 熟考人間乱数 180 ファイルの学習マップ

Fig. 3 SOM map of the 180 'deliberate' data

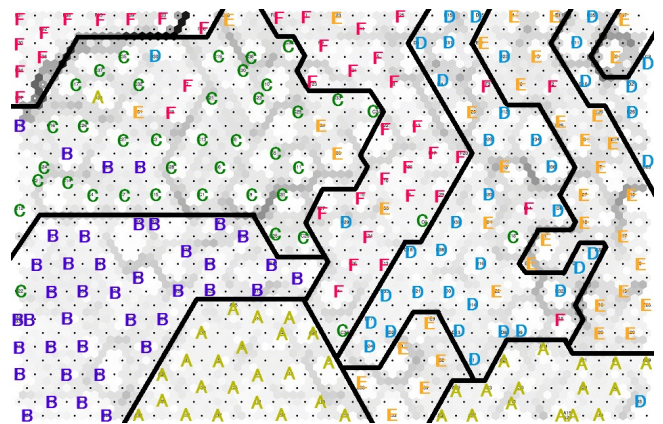


図 4 境界線を引いた直感人間乱数の学習マップ

Fig. 4 SOM map of 'intuition' data with the borderlines

しなかった直感人間乱数一人各 10 ファイルを用いた。識別用のデータを図 2 上の学習マップに入力し、図 4 の境界線を引いた結果を図 6 に示す。図 6 を見ると概ね同一人物の領域に分類されているのが分かる。自分の領域に分類された各 10 ファイルの割合を認識率とし表 7 にまとめた。表 7 を見ると A, D, F は 9 割以上, B, C, E は 6, 7 割が自分の領域に分類されている。また、図 6 を見ると C, E の一部は

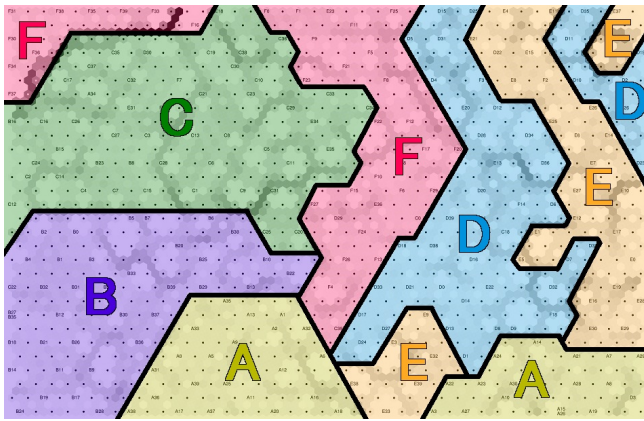


図 5 各個人の領域

Fig. 5 Territories of individual subjects

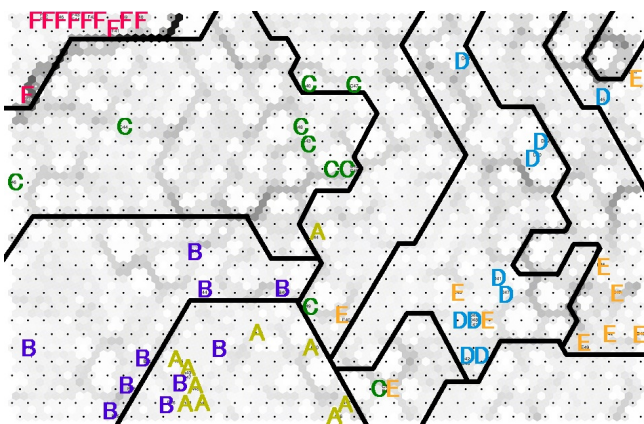


図 6 テストデータの分類結果

Fig. 6 SOM map of test data (open test)

表 7 図 6 における認識率

Table 7 The recognition rate as in Fig. 6

| 被験者 | A | B | C | D | E | F | Ave |
|---------|----|----|----|-----|----|-----|-----|
| 認識率 (%) | 90 | 70 | 60 | 100 | 60 | 100 | 80 |

境界線を挟んで隣に分類されており概ね正しく分類されているのが分かる。

6. おわりに

本研究では直感データと熟考データを比較することで人間乱数に現れる個人の特徴に変化があるかどうかを探ると共に、SOMを利用した人間乱数データによる個人識別の可能性についての研究を行った。その結果、直感データと熟考データのいずれにおいても個人を明確に分類する指標を特定することができなかつた。しかし、H,ADJ,TPL,PL,RG,RP,TKD,TIMEの8個の指標を特徴量として同時に使い、SOMによる教師なし学習を適用することにより、直感人間乱数データを概ね個人別に分類することに成功した。熟考人間乱数データについては、分類できた個人も見られたが複雑に混ざり合っている部分も多く、対象とした全ての個人別には分類できなかった。これ

により直感的に生成した人間乱数による個人識別の可能性を追求することにした。しかしその結果は、直感データに対しても認識率は平均80%程度であり、低い人は60%程度に留まるなど、明確な分類が出来なかつたデータも多くあつた。これは個人内のデータにばらつきが大きいために、今回使用した指標では個人内と個人間の差を明確に分けることができなかったことが原因である。また、データの採取日にバラつきがあるなど、データ採取環境が整っていなかつたことも個人内誤差の原因と考えられる。しかし、TKD,TIMEの2つの指標については個人内誤差が少ない傾向が見られた。この2つの指標は、他の6つの指標が乱数列の特徴を見るのに対し、被験者の乱数生成時の行動を見るものであつた。このことより人間乱数から個人の癖を捉えるには生成された数列の解析と同時に、生成時の被験者の行動を解析することが重要であると考えた。以上のことより人間乱数を個人認証システムに利用するためには数列と行動2つの特徴を組み合わせることが有効であるとする。そのため今後の課題として、生成方法や採取環境を整備することで個人内のばらつきを減らし、またキー入力タイミング等生成時の行動を新たな特徴とすることで個人識別の精度を高めることが今後の課題である。

参考文献

- [1] 乱数テスト研究会 (1973), “人間乱数—頭脳のプリズム—”, 自然, 8月号 (中央公論社), pp.49-57.
- [2] W.A.Wagenaar(1972), “Generation of Random Sequences by Human Subjects: A Critical Survey of Literature”, Psychological Bulletin, Vol.77, pp.65-72.
- [3] 楊静宏, 川原正弘, 五味壮平, 新貝御蔵 (2006), “人間乱数の分析”, The 20th Annual Conference of the Japanese Society for Artificial Intelligence, 1A1-2.
- [4] 矢内浩文, 森太香夫 (2005), “人間が生成するランダム系列の性質—シンボルの種類と生成手段への依存—”, 電子情報通信学会技術研究報告, NLP2005-71.
- [5] 森太香夫, 矢内浩文 (2005), “人間が生成するランダム系列の性質(その2)—数字・かな・英字, そして口頭・キーボード—”, 電子情報通信学会技術研究報告, NLP-85.
- [6] 三島雅史, 田中美栄子 (2007), “短い人間乱数による診断可能性と指標の選定”, 情報処理学会論文誌: 数理モデル化と応用 (TOM19), pp.47-54.
- [7] 田中美絵子, 榎本良太 (2008), “逆テンキー (MPK) 方式による短い人間乱数”, 情報処理学会研究報告 (2008-MPS-88), pp.27-30.
- [8] J.N.Towse, D.Nell(1998), “Analyzing Human Random Generation Behavior: A Review of Methods Used and a Computer Program for Describing Performance”, Instruments & Computers, Vol.30, pp.583-591.
- [9] 徳高平蔵, 岸田悟, 藤村喜久郎 (1999), “自己組織化マップの応用”, 海文堂.
- [10] T.コホネン (著), 徳高平蔵, 大藪又茂, 堀尾恵一, 藤村喜久郎, 大北正昭 (監修)(2005), “自己組織化マップ 改訂版”, シュプリンガー・ジャパン.
- [11] 大北正昭, 徳高平蔵, 藤村喜久郎, 権田英功 (2008), “自己組織化マップとそのツール”, シュプリンガー・ジャパン.
- [12] SOM-PAK, <http://www.cis.hut.fi/research/som-research/nmrc-programs.shtm1>