

逆テンキー (MPK) 方式による短い人間乱数

田中美栄子[†], 榎本良太[‡]

鳥取大学大学院工学研究科情報エレクトロニクス専攻[†], 富士通テン (株) [‡]

概要

人間乱数テストはヒトの乱数発生に現れる特徴から思考力や集中力のレベルを判定し、精神的疲労度の測定などを行おうとするもので、筆者等も数年来、健常者の人間乱数に表れる僅かな差異を認識する方法について、隠れマルコフモデル (HMM) や相関次元の測定などの方法で研究してきた。しかしこれらは長い乱数列を必要とし、被験者に対する負荷が大きい事が難点である。そこで長さ 50 の短い乱数列から簡便に特徴を取り出せる指標を探した結果、RP, ADJ, TPI, H の 4 指標の組み合わせが有効であることを見出し 2006 年の研究会で発表した。今回は更に短い列を用いる目的で、携帯電話の入力画面を使ってテンキーの数字を 1 個ずつランダムな順に打つ、という形のデータ採取を行い (これを逆テンキー方式の人間乱数と呼ぶ)、その有効性を考察した。

Human Random Generation Test by Using the Mobile Phone Keyboard

Mieko Tanaka-Yamawaki[†], Ryota Enomoto[‡]

Department of Information and Electronics, Graduate School of Engineering, Tottori University[†], Fujitsu Ten Co. [‡]

Abstract

Human Random Generation is a psychological test by which characteristics of individual person, or degree of mental fatigue can be measured. We have been practicing over several years including the time series analysis by means of hidden Markov model (HMM) and correlation dimensions. More recently, we focused on developing a method of extracting characteristics of very short sequences identified a set of indices suitable to characterize short sequences. In this report, we propose a novel direction of human random generation test using the number keys (Ten-Key) of mobile phones, by which we fix the length of data sequence to 9 and measure the randomness of the order of 9 number keys (1 - 9) on the mobile phone keyboard (MPK).

1. はじめに

近年、益々の加速を続ける高齢化社会において、加齢による記憶力低下や認知症などの脳障害は大きな問題を生じ、社会的関心事になっている。認知症などの一番の解決策は早期発見であり、症状の出始める時期にそれを知ることができれば非常に有効なものになる。一般的に使われている自己診断方法としては、例えば「今日は何曜日ですか?」、「昨日の天気はどうでしたか?」と言ったような直接記憶を辿る質問形式が多く、症状がある程度進行していない限り間違えようのない問題が出題される事が多い。しかし早期発見のためには、本人も自覚症状がないような初期段階での早期発見が望ましいわけで、我々はこの目的に人間乱数テストが利用できそうだと考えから、人間乱数の診断への応用可能性を数年にわたって探ってきた。[1-5]

人間乱数 [6-7] とは、人間に乱数生成を指示して得られた乱数列のことを指し、被験者の癖や生成の際の心理状況、脳の状態などを反映する。データ採取の際に特別な装置を必要とせず、被験者はただ数字を言うだけであるから、毎日行うことができ身体的な負担も軽微であることから非侵襲診断として期待され早期発見に繋がると考えられている。しかしただ早期発見を行うだけでなく、症状の進行の抑制、改善を行えるような方法があれば更に有効である。

乱数生成の際にはかなりの「集中力」を要求することから、その側面を積極的に利用することを考えた。漠然と数字を生成するのではなく、被験者に、より乱数生成に意識を向けさせることで、集中力アップに繋がるようなやり方が望ましい。例えば「百マス計算」は単純な加減算を行うものであるが、できるだけ正確

にできるだけ速く解くことを何度も繰り返すことで計算力向上を図る。また任天堂 DS ソフトの「脳トレ」もその一例である。単純な問題が多いが脳年齢などの診断結果を導きだしてくれる。どちらにも共通しているのは、ある程度の「集中力」を要求することと、主目的として診断よりもトレーニングを図ろうとしていることである。

そこで、本研究では、人間乱数を医療分野、教育分野などの他分野への適用を考え、新たに携帯電話の1～9のキーを用いた、1から9までの数字のランダム列への並び替えを被験者に要求する人間乱数テストを考案し（これを「逆テンキー方式の乱数生成」と呼ぶ）、その有効性を確認するための基礎研究として、個人の特徴抽出および乱数度測定を行った結果を報告する。

2. 逆テンキー方式乱数生成

携帯電話の数字画面は1-2-3が上段に、4-5-6がその下に、7-8-9が更にその下に並び、いわゆるテンキーとは上下が逆になっているので、この画面を逆テンキー画面と呼び、この3×3の升目に並んだ9つの数字を一回ずつランダムな順番に入力する形での乱数生成を、逆テンキー方式の人間乱数生成と呼ぶことにする。逆テンキー方式の利点は、データ長を短くでき、しかも一定長に保てることと、データ生成時間を短くすることによる被験者の疲労度の軽減と共に、集中力の訓練効果も同時に期待できることである。

データ採取には携帯電話の1～9のキーを使用し、以下の方法で乱数データを採取した。

- 1) 1～9のキーを1回ずつ押す(1試行)を3回繰り返す(3試行 = 1ファイル = 27個)。
- 2) 打つ際に液晶画面は不可視(キー配置は可視)
- 3) キーを押すのは片手のみ
- 4) 乱数生成の時間は被験者に任せる

個人の特徴抽出は、1試行(9個)毎に行うが、乱数度の測定にはもう少し長い列が望ましいので、1ファイル(27個)毎に行うこととする。2)で液晶画面を不可視にしたのは、前回の試行で生成したデータ列を目で見ると、今の試行列に対して意図的に違う乱数列にしようとする余計な思考が働くのでこれを防ぐためである。あくまでも、頭の中で数字を思い浮かべ、できるだけランダムになるようにキーを押してもらい、データ採取を行った。3)で片手とするのは、キーからキーへの移動距離をデータ解析に用いるので、両手を用い

ると正確な距離データが採取できないためである。

本研究では、A～Gの7人(21～23歳までの男性を被験者とした。1回のデータ採取は27個(1ファイル)ずつとし、被験者1人当たり計40回(27個×40ファイル分)行った。

個人の特徴抽出には、これらの3指標ニューラルネットワークを用いて、1試行ずつ(9個単位)で解析を行い、採取した40ファイル(9個×120試行)中、30ファイル(9個×90試行)を学習用データ、10ファイル(9個×30試行)をテスト用データとして使用した。

また、乱数度の測定にも、同様のデータを使用した。データ採取前に予めどういった指標を用いて乱数度を決定するのかを説明し、解析は1ファイル毎(27個単位)で行った。これは乱数度の高低についての基準が被験者ごとにばらつくのをできるだけ避けようとの意図に基づくものである。

3. 特徴抽出の方法とその結果

パターン抽出には<領域分布>と<移動距離>の組合せを使用する。<領域分布>は、9個ずつあるデータ列を左から3個ずつ、左、中間、右領域の3領域に分け、それぞれのキー数値の出現位置を調べ、その偏りから特徴を抽出する。例えば図1のデータに対し、「1」、「3」、「5」、「8」は各領域に1回ずつ出現し偏りが見られないのに対し、「2」は左領域に3回出現し、それ以外の領域には出現しないという偏りを持つ。そこで各数字の領域分布をラベル化するため、左領域にある数字を全てIに、中間領域にある数字を全てIIに、右領域にある数字を全てIIIにし、パターン認識においては元データをI, II, IIIの列に変換して使用した。

1試行目	1 2 3	4 5 6	7 8 9
2試行目	2 5 6	8 4 3	9 1 7
3試行目	8 2 9	6 7 1	5 4 3
	左領域	中間領域	右領域

図1 データ列の3領域区分け例

Fig.1 Image of the 3 Regions in a Data Sequence

<移動距離>は各々のキーを3×3の升目に対応する座標に対応させ(図2)、キーからキーへの移動距離の総和を算出する。距離算出に関しては、小数点第2位以下は切り捨てる。例えば、「1」から「2」への移動距離は1、「5」から「3」へは2、「3」から「7」へは2の平方根の2倍で、2.8となる。

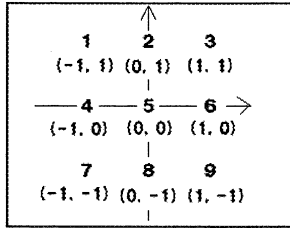


図2 キー数値の座標対応

Fig.2 Cartesian coordinates of the Keys

1 試行あたりの移動距離(d)は、 $8 \leq d \leq 17$ の範囲の値を取る。これを表2のように $d=11.4, 13, 15$ の3点で4領域に分けて小さい順に0, 1, 2, 3でラベル付けする。例えば、 $x = 256843917$ というデータに対しては、 $d=1+1+1.4+1.4+2.2+2+2.8+2=13.8$ となるので、移動距離は2とラベル付けされる。

4. ニューラルネットワーク (NN) による学習

特徴抽出には3層 NN を用いて BP 学習を行う。入力層には10素子、中間層には3素子、出力層には1素子を配置する。採取したデータ列をそのまま NN に入力してもあまり良い分類結果は得られないため、1~9の数字が出現した領域を I~III でラベル化して NN の最初の9入力素子への入力とし、10番目の入力素子にはそのデータ列の〈移動距離〉ラベルを入力する。学習は10000 回行った。パラメータ(結合荷重Wおよび閾値 θ)の初期値には0~1の乱数を与え、被験者ごとに適当な整数値の教師信号を与え、学習用データ 30 ファイルを学習させた後、テスト用データ 10 ファイルを使用して個人の認識率を算出した。学習に用いた NN の構造を図3に示す。入力層に10素子すべてを用いる場合と、10番目の〈距離分布〉を除いた〈領域分布〉のみの情報からなる場合とを比較して表1に示す。

〈領域分布〉のみの場合より、〈領域分布〉と〈移動距離〉の両方を用いた場合の方が、いずれの被験者に対しても認識率の高い事がわかる。また、被験者Dのみが33%と低い認識率であるものの、被験者A, B, Cは90%を超える高い認識率を示し、全体の平均でも80%の認識率を得た。これより、本稿の方法で特徴抽出が成功して

いると言える。

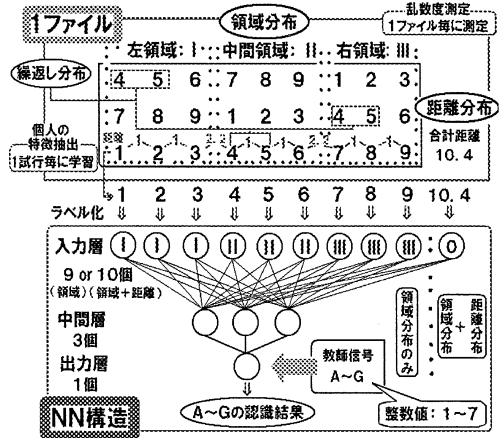


図3 学習に用いたニューラルネットワーク

Fig.3 Neural network used for neural network

表1 個人認識率の比較

Table 1 Comparison of Recognition Rates

被験者	A	B	C	D	E	F	G	平均
領域のみ	90	73	53	0	57	53	60	55
領域+距離	100	93	97	33	70	80	90	80

4. 乱数度

乱数度(R)は3試行分、27数字を対象に計測する。Rはデータに特徴がなければ大きく、偏りの目立つ場合に小さいと考えられる。そこでRの反復度、領域分布、移動距離を用いた定義を考える。

反復度は文献[1]で提案したりピート・パターン(RP)と同様に、隣接する2数字の反復度を示すものであり、ここではそのRへの寄与分 R_{RP} として、

$$R_{RP} = 24 - r \quad (1)$$

を採用する。これは「反復度が少ないほど、そのデータ列には特徴がない」との考えから、反復の起きる隣接2文字列の数rをその最大数24から差し引いたものとしている。最大数が24である理由は次のようである。図1に示すように、長さ9の数字列を3列取っているため、例えば「1-2」というパターンが3列全てに

出現し、それ以外には3領域のいずれにも反復が起きないとすれば $r=3$ となる。一方、3列が全く同じ数列の場合には8種類の隣接数字の組が各々3回出現することになるので、 $r=24$ がここで定義する最大数と考えるのである。

領域分布の R への寄与分 R_{PT} については、図1のようなデータ構造において、出現位置の偏りが最も小さくなるのは、各領域に1~9の数値が1回ずつ使用された場合である。そこで n を各領域に1回ずつ偏りなく出現する文字数としてそれに重率2を掛けて

$$R_{PT}=2n \quad (0 \leq R_{PT}=18) \quad (2)$$

とする。

移動距離の R への寄与分 R_D は、列 i における移動距離ラベル d_i を3列分足し合わせたものとする。 d_i は0から3までの整数で定義したので3領域共に最大値をとる場合に R_D の最大値は9となる。

$$R_D = \sum_{i=1}^3 d_i \quad (0 \leq R_D = 9) \quad (3)$$

以上を総合して、乱数度(R)を次式で定義する。

$$R = (R_{RP} + R_{PT} + R_D) / R_{MAX} \quad (4)$$

乱数データ列から乱数度を測定する例を図4に示す。

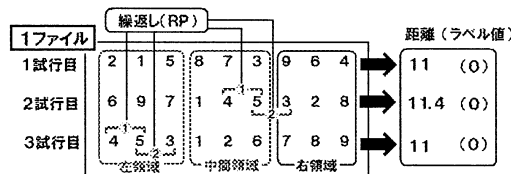


図4 乱数度測定データ列の例

Fig.4 Example of computing randomness (R)

こうして測定した乱数度と前章までの個人認識率との関係を被験者ごとに比較した結果を図5に示す。被験者Dは認識率が低く、乱数度が高い。これは、各ファイル間の関連性が弱く、生成する乱数列にも特徴がないことを示すと考えられる。逆に被験者Aと被験者Gは、各データ間の関連性が強く、同じような特徴を持った乱数列を生成している。

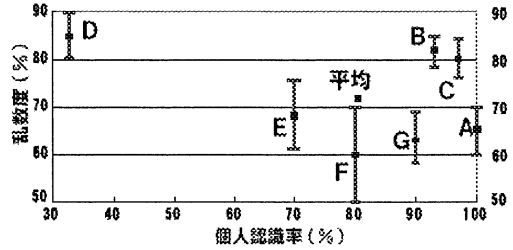


図5 各被験者に対する乱数度と認識率の関連性

Fig.5 Relationship between Randomness and Recognition Rate for individual Subject.

5. まとめ

本研究では、新たに携帯電話を用いた逆テンキー方式を採用した。その結果、従来よりも、少ないデータ列で、ある程度の個人の特徴抽出ができ、乱数度との関係も見出すことが、人間乱数の新たな可能性を見出すことができた。

参考文献

- [1] 三島雅史, 田中美栄子, “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 Vol.48 No.SIG 19 (TOM 19), pp.47-54(2007)
- [2] 田中美栄子, 川越美姫子, 増田博子, “人間乱数と脳機能モデル” 宮崎大学工学部紀要, pp. 297-301 (1998)
- [3] Tanaka-Yamawaki M (1999), Human Generated Random Numbers and a Model of the Human Brain Functions; Proc.1999 IEEE SMC, pp. 223-228
- [4] 田中美栄子, “人間乱数の特性解析とその応用” 平成19年度統数研乱数重点型共同研究第2回研究会「乱数の応用指向特性評価とその応用」(2008年1月4日, 統計数理研究所)
- [5] 乱数テスト研究会, “人間乱数 -頭脳のプリズム-”, 自然 1973年8月号(中央公論社), pp.49-57(1973)
- [6] 橋文夫 “人間による乱数列をマルコフ連鎖としてみたときの特性” 情報処理学会論文誌 Vol.20, No.1(19790115) pp. 1-7