

ACT-R を用いた記憶エラーのモデル化による人間の内部状態推定 Internal State Estimation According to Modeling of Memory Error Using ACT-R

新堀 耕平[†], 西川純平[‡], 長島 一真[‡], 森田 純哉[§]

Kohei Shimbori, Jumpei Nishikawa, Kazuma Nagashima, Junya Morita

[†]静岡大学情報学部, [‡]静岡大学創造科学技術大学院, [§]静岡大学大学院情報学領域

Faculty of Informatics, Graduate School of Science and Technology, College of Informatics, Shizuoka University
shimbori.kohei.21@shizuoka.ac.jp

概要

記憶のエラーは個人属性や感情状態など、様々な要因により生じる。エラーは、想起されるべきでない情報が思い出されるコミッションエラー、想起すべき情報を思い出すことができないオMISSIONエラーの2種類に大別される。本研究では、これらのエラーをクラウドソーシングにより取得し、それらと ACT-R モデルとの整合を検討した。結果、いくつかの感情評価項目とモデルパラメータとの間に相関がみられ、限定された環境において、人間の個人傾向を推定可能であることが示された。

1. 背景

人間の記憶は、過去と現在、現在と過去を結び、個人のアイデンティティの基盤となる。しかし、人間の記憶は日常生活の様々な場面においてエラーを引き起こす [1]。これらのエラーには思い出すべきではない記憶を思い出してしまうコミッションエラー、思い出すべき記憶を思い出すことのできないオMISSIONエラーが含まれる。どのようなエラーがどのような状況において生じるのかは、年齢や成育歴などの個人の属性、あるいはそのときどきの個人の体調や感情によって変化すると想定できる。この想定を換言すれば、表出されたエラーから個人の内部状態を推定することも可能ともいえる。本研究では、上記の可能性（表出された想起エラーからの内部状態推定）を、認知モデルの利用により検討する。この目的を達成するために、多量の想起エラーを簡便な手法により収集する実験を実施する。そして、その実験により得られたデータに対して、認知アーキテクチャである Adaptive Control of Thought-Rational (ACT-R) [2] により構築されたモデルを利用したパラメータフィッティングを行う。ACT-R は汎用的な認知アーキテクチャであり、多様な個人の多様な状況における認知プロセスを説明するパラメータセットを有している。個人の想起エラーとフィットするパラメータを組織的な探索によって抽出したのちに、抽出されたパラメータが個人の属性や感情状態と

対応付けられることを検討する。

なお、本研究は著者らによる過去の発表 [3] で構築されたモデルをベースに、よりデータと適合するパラメータを探索したものである。以下、本研究で得られた結果を説明するために、既発表の実験とモデルを示す。

2. 実験

数列の記憶課題を採用した。本研究の課題において、参加者はモニターに提示される 10 桁の数列を記憶し、報告することを求められる。この際、参加者は「すべてを正確に覚えることが困難であったとしても、回答時には可能な限り頭に残っている数字を復元し、回答欄に記入するに努力」をするよう教示される。提示される数列はランダムに生成され、提示時間は 2 秒間である。数列の提示が終了した後、20 秒間の回答時間内に、モニター上のテキストボックスに記憶した数列をキーボードから入力する。1 つの数列の回答が終了した段階で、モニター上のボタンやエンターキーを押すことによって即座に次の数列の試行に移行できるものとした。

実験には、クラウドソーシングにより募集された平均年齢 43 歳の 50 人が参加した、参加者は 30 個の数列について記憶課題を実施した。記憶課題終了後、Google フォームによって作成されたアンケートに回答した。アンケートには、個人の属性に関する項目、体調、暗記力自己評価、日本語版 PANAS [4] に含まれる感情評価 20 項目が含まれた。

実験環境の不備により正確な回答を得られなかった試行を除外し、それ以外のものを有効な回答とした。課題数列と回答数列の類似度は、レーベンシュタイン距離（編集距離）とした。2 つの数列が完全に一致した場合には 0 となる。

この実験により得られた参加者全員の回答のヒストグラムは図 1 のようになった。

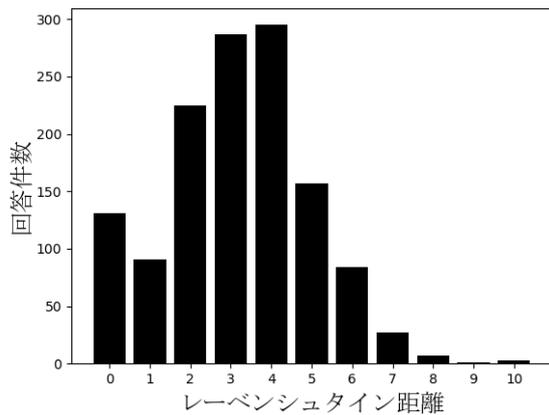


図 1 実験結果

3. モデル

実験によって得られたコミッションエラー及びオMISSIONエラーを再現するために、ACT-R Tutorial Unit5 [5] に含まれる grouped モデルを一部改変して使用した。このモデルは 1234567890 という単純な数列の再生をシミュレーションするものである。この数列を、モデルは (123)(456)(7890) のようにグループを形成することでエンコードしている。各グループ内で、個々の数字はグループ内の位置 (1 番目から 4 番目) と紐づけられて記憶される。

数列の想起において、モデルは数列の左から順に、グループ及びグループ内の位置をシフトしながら数字を報告する。すなわち、課題の初めにおいて、モデルは「1 番目のグループ」の「1 番目の数字」を記憶から引き出すことを試みる。次に、2 番目、3 番目の数字を検索し、位置に対応する数字の記憶を引き出せなかったとき、「2 番目のグループ」の数字の検索に移行する。これを「3 番目のグループ」の「4 番目の数字」まで繰り返して再生する。

本研究のモデルにおいて、記憶内の数字 i には以下の式 1 によって定義される活性値 (A_i) が付与される。

$$A_i = \sum_l PM_{li} + \varepsilon \quad (1)$$

モデルは、数列の各位置の数字を検索する際に、記憶内に含まれるすべての数列内の数字に対して活性値を計算し、最も活性値の高い数字を検索結果として出力する。式 1 において、 l は検索リクエストが含む条件の数であり、本モデルでは、検索対象となる数字のグループ、グループ内でのチャンクの場所が検索条件となる。そのため条件の数 l は 2 となる。 M_{li} は検索の条

件と検索対象となる数字 i の対応する属性との類似度を示す。両者が完全に適合したときには 0.0、適合しない場合には -1.0 が設定される。両者に何らかの相似がある場合には、その度合いに応じて 0 から -1 の値が割り振られる。本モデルでは、隣接するグループの組 (1 番目のグループと 2 番目のグループ、2 番目のグループと 3 番目のグループ)、グループ内の隣接する位置 (1 つ目の位置と 2 つ目の位置、2 つ目の位置と 3 つ目の位置、3 つ目の位置と 4 つ目の位置) について、-0.5 を設定する。

類似度は式 1 において P によって重みづけられる。この係数は ACT-R の内部では、グローバルパラメータ:mp (mismatch penalty) として設定される。:mp の値が大きき場合と比較して、この値が小さき場合には、検索リクエストが本来求めている数字が有する活性値と検索リクエストとは完全には適合しないものの類似する属性を持つ数字の活性値の差分は小さくなる。

そして、類似する数字の活性値は、活性ノイズ (ε) の値により、ときとして、本来求めていた数字の活性値を上回る。 ε はロジスティクス分布に従う確率変数である。分布の幅は ACT-R のグローバルパラメータ:ans (activation noise s) によって設定される。:mp が小さく設定され、:ans が大きく設定されたときには、検索リクエストが本来求めない数字が検索されるコミッションエラーの確率が高くなる。本研究のモデルにおいて、そのようなエラーは数字の順序が入れ替わるエラーとして観察されることになる。

一方、いずれの記憶も思い出されないオMISSIONエラーは、記憶内のすべての数字の活性値が設定された閾値を下回ったときに生じる。ACT-R において、閾値はパラメータ:rt (retrieval threshold) によって設定される。このエラーは数字が報告されないという形で観察される。

ACT-R grouped モデルにおいて設定されるデフォルトのパラメータ値を用い、モデルを 1000 回実行した。各試行結果に対して、モデルが本来想起すべき数列 (1234567890) とモデルが出力した数列のレーベンシュタイン距離を計算したところ、図 2 のようなヒストグラムが得られた。図 1 と図 2 を比較すれば、モデルは人間に対して、本来想起すべき数列と完全に一致する数列 (距離 0) を多く出力していることがわかる。この結果の差異は、数列の記憶に関わる設定の差に由来すると考える。本研究のモデルには数列の記憶に関わるプロセスが含まれておらず、モデルは事前に記憶に格

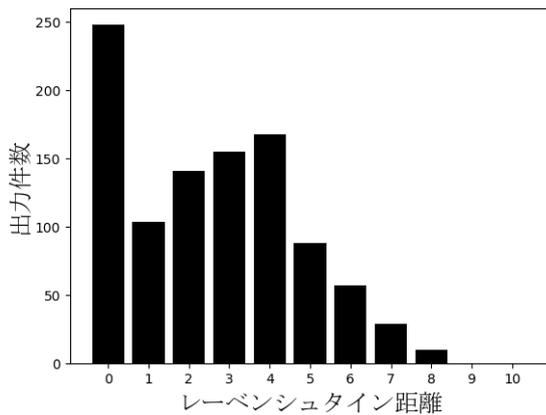


図 2 パラメータのデフォルト値による結果

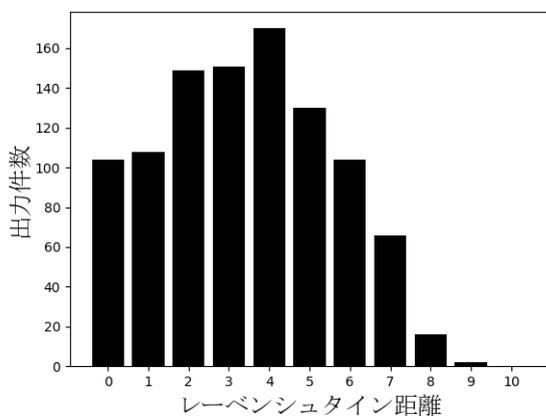


図 3 パラメータフィッティングの結果

納された数列 (1234567890) を想起した。それに対して、人間は 2 秒間という短時間の間に行われた不完全な記憶の条件で記憶を想起した。

このような差異があるものの、モデルの結果はレーベンシュタイン距離が中間の記憶を想起するという、人間の結果 (図 1 を参照) と類似する部分もある。さらに、先述のとおり ACT-R にはコミッションエラーやオMISSIONエラーの生起確率を変化させるパラメータが多く存在する。よって、それらを調整することにより、データとモデルを対応付けることは可能と考えられる。これをもとに、ACT-R のパラメータを調整し、データフィッティングを行った結果について次節にまとめる。

4. 結果

初めに、1 節において述べた可能性 (表出された想起

エラーからの人間の内部状態推定) を検討するために、2 節の実験において得られた全回答 1308 件のデータに対し、3 節のモデルパラメータを探索することで、フィッティングを試みた。

本研究での探索対象となったパラメータは、コミッションエラーに関わる:mp (0.2~2.0, 0.2 刻み)、オMISSIONエラーに関わる:rt (-3.0~1.0, 0.5 刻み)である。パラメータの探索範囲は、先行研究 [3] で採用した設定をベースとしつつ、実行時間の制約を考慮して定めた。実験データとモデルの適合を測定するために、上記のパラメータの組み合わせからなる 100 の ACT-R モデルについて、図 2 と同様のヒストグラムを構成した。そして、図 1 のヒストグラムに対して、最も適合するモデルのヒストグラムを、ヒストグラムインターセクションにより評価した。

結果として図 3 のヒストグラムが得られた。ヒストグラムインターセクションの値は 0.981¹、ヒストグラム間での相関係数の値は 0.952 と高い適合が得られた。この時の:mp の値は 0.8、:rt の値は -0.5 であった。フィッティング前と比較し距離 0 の回答件数は減少しているが、実験結果のヒストグラムと比較して距離 6 以上のものが多くなっている。

次に、実験で得られた個人のデータに対して、同様にフィッティングを行った。各参加者が報告した 30 の回答に対して課題となった数列とのレーベンシュタイン距離を計算し、ヒストグラムを構成した。そして、得られた各被験者のヒストグラムと最もヒストグラムインターセクションが大きくするモデルのパラメータを探索した。

表 1 は上記の処理により出力されたモデルパラメータとアンケートによって各個人から得られた回答とのスピアマンの順位相関係数を示している。なお、表中の HI は各個人に対して見積もられたヒストグラムインターセクションとアンケート項目の相関を示している。今回の 100 のモデルから十分に適合するモデルが見つからなかった参加者も存在する。そのような参加者の特徴を検討するために、HI と参加者の属性や内部状態との相関も検討した。表より、モデルの評価指標、出力されたモデルパラメータに対し、それぞれ有意となる属性や感情評価項目がわかる。HI に関しては、年齢との間で負の相関が得られており、また「恐れた」

¹ 実験データ同士のヒストグラムインターセクションを最大値として正規化。

表 1 個人モデルとアンケート回答の相関

	HI	mp	rt
年齢	-0.366*	0.114	-0.226
性別	-0.136	0.183	-0.122
体調	0.051	0.097	0.019
最終学歴	-0.145	0.140	-0.081
暗記力自己評価	-0.001	-0.046	0.233
強気な	-0.168	0.222	0.023
やる気がわいた	-0.032	-0.024	-0.092
活気のある	-0.081	-0.053	-0.042
熱狂した	-0.025	0.082	-0.049
興味のある	0.032	0.135	-0.085
興奮した	0.049	0.097	-0.030
誇らしい	-0.008	0.099	-0.023
機敏な	0.010	0.106	-0.111
決心した	0.039	0.134	-0.202
注意深い	0.080	-0.184	0.066
ポジティブ	0.017	0.083	-0.048
恐れた	0.286*	-0.221	0.103
おびえた	0.195	-0.182	0.023
うろたえた	0.259	-0.243	0.004
恥ずかしい	0.103	-0.217	-0.056
うしろめたい	0.045	-0.029	-0.244
びりびりした	0.027	-0.107	0.000
苦悩した	0.180	0.058	-0.134
イライラした	0.140	-0.078	-0.307*
神経質な	0.109	-0.176	-0.166
敵意を持った	0.051	-0.133	-0.061
ネガティブ	0.187	-0.178	-0.095

*は無相関検定の結果 $p < .05$

という感情評価項目に対して正の相関が得られている。これより、今回のパラメータ設定の範囲においては、年齢が低い参加者、また「恐れた」と感じた人の回答結果によく適合するモデルが得られたと考えられる。また、 rt と「イライラした」という感情評価項目に対して負の相関が得られた。 rt が低いほど、より活性化の低い数字が想起される確率が増加し、モデルは数字

の想起に失敗しなくなる。つまり、表 1 に観察される rt と「イライラした」の負の相関は、苛立ちを強く感じた人ほど、オMISSIONエラーを起こしづらかったことを示す。課題中の苛立ちを、記憶の想起がうまく行かないことに対して生じた感情、あるいは課題への動機づけを低める感情と考えるのであれば、この結果の解釈は困難である。しかし、実験後にイライラしたと報告したということは、課題に対してより辛抱強い態度で臨んだものと考えられる。そして、そのようなより集中した課題の遂行が、数列の正確な記憶を達成したと考えることは理にかなっている。

5. まとめ

本研究では記憶のエラーから、その原因となる内部状態を認知モデルとのフィッティングにより推定できるかを検討した。記憶のエラーを誘発する実験で得られたデータに対して、組織的なパラメータフィッティングを、全体及び個人のデータに対して実施した。その結果、一部の感情評価項目とモデルパラメータ間に有意な相関が観察された。しかしながら、前節の解釈に示されるように、今回報告された相関は、本研究の実験状況に限定されている可能性がある。より踏み込んで言えば、記憶検索時の閾値は、特定の感情状態と直接対応するパラメータというより、その原因となる個人傾向（辛抱強さなどに対応）と考えられる。個人の内部状態と直接対応するパラメータを検討するためには、より多様な状況での実験を実施する必要があると考える。また、本研究において得られた結果は、同じデータを対象として実施した先行研究 [3] のものと異なっている。よって、今後フィッティングの精度を上げ、より個人にフィットしたモデル化を行うことで、内部状態の推定可能性が高まると予想する。

文献

- [1] Schacter, D. L. (2001) "The Seven Sins of Memory: How the Mind Forgets and Remembers"
- [2] Anderson, J. R. (2007) "How can the human mind occur in the physical universe?"
- [3] 新堀耕平, 西川純平, 長島一真, 森田純哉, (2023) "認知モデルを用いた想起エラーにおける人間の状態推定", HAI シンポジウム 2023
- [4] 川人潤子, 大塚泰正, 甲斐田幸佐, 中田光紀, (2012) "日本語版 The Positive and Negative Affect Schedule (PANAS) 20 項目の信頼性と妥当性の検討", 広島大学心理学研究
- [5] Bothell, D. (2022) "Unit5: Activation and Context", ACT-R tutorial (Version 7.27.7)