

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

Cognitive Modeling for Estimating Human Internal States in Memory Recollection Errors

新堀耕平¹ 西川純平² 長島一真² 森田純哉³

Kohei Shimbori¹, Jumpei Nishikawa², Kazuma Nagashima² and Junya Morita³

¹ 静岡大学情報学部

¹ Faculty of Informatics, Shizuoka University

² 静岡大学創造科学技術大学院

² Graduate School of Science and Technology, Shizuoka University

³ 静岡大学学術院情報学領域

³ College of Informatics, Shizuoka University

Abstract: 人間の記憶のエラーは多用な原因により生じる。代表的なエラーの原因として、想起すべき際に情報を思い出すことができないオMISSIONエラー、想起されるべきではない情報が呼び出されるCOMMISSIONエラーが想定される。人間により生成された記憶のエラーから、これらの原因を推定することは一般的には困難である。本研究では、これらのエラーに関する多用なサンプルをクラウドソーシングにより取得し、取得されたエラーと ACT-R モデルとの整合を検討することで、各エラーの原因を分析する。

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ふ COMMISSION エラー、思い出すべき記憶を思い出すことのできない OMISSION 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。

2 実験

2.1 方法

2.1.1 課題

数列の記憶課題を採用した。単純であるため、モデル化が容易であること、提示桁数や提示時間、提示間隔を操作することで、難易度を調整可能であることを理由とした。

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、報告することが求められる。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のちに、20 秒間の回答時間のなかで、画面上のテキストボックスに記憶した数列を入力する。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」をするように指示

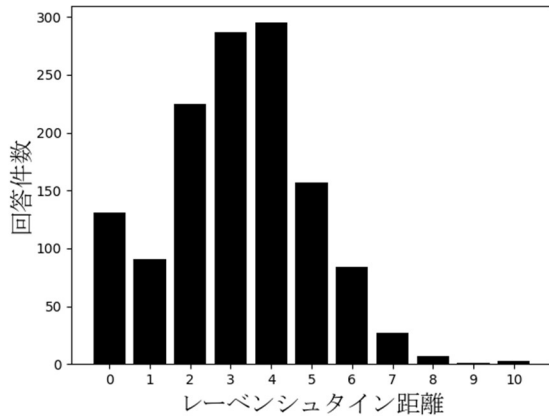


図1 実験結果

される。なお、1つの数列に対する回答が終了した段階で、画面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。

2.1.2 参加者

クラウドソーシングにより募集された50人(男性37名、女性13名)が参加した。参加者の平均年齢は43.36歳(27~66歳, SD≒8.35)であった。参加者は、以下の手続きを遂行することにより100円を受け取った。

2.1.3 手続き

参加者は30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、回答者個人の属性(年齢、最終学歴)に関する項目、暗記力自己評価(1:とても悪い~5:とても良い)体調の報告(1:悪い~5:良い)、日本語版PANAS[3]に含まれる感情評価20項目(ポジティブ感情に関わる項目10、ネガティブ感情に関わる項目10)が含まれた。感情評価の各項目については6段階(6:非常に当てはまる~1:全く当てはまらない)で評価を収集した。

2.2 結果

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。有効な回答件数1308件のうち、正答は131件、誤答は1174件、無回答は3件であった。また、誤答のうち10桁のものは601件、10桁未満のものは522件、10桁を超えるものは51件であった。

また、課題数列と回答数列との一致については、レーベンシュタイン距離(編集距離)によって計測した。2つの数列が完全に一致した場合には0、共通

点が存在しない場合は10となる。実験によって得られた回答のヒストグラムは図1のようになった。

3 モデル

3.1 数列の表現および検索プロセス

実験によって得られたコミッションエラーとオMISSIONエラーを再現するために、ACT-R Tutorial Unit5 [4]に含まれるgroupedモデルを一部改変した。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。各グループの中で、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、グループおよびグループ内の位置をシフトしながら数字を報告していく。すなわち、課題のはじめにおいて、モデルは「1番目のグループ」の「1つ目の位置の数字」を記憶から引き出すことを試みる。その後、「1番目のグループ」の「2つ目の位置の数字」の検索がなされ、「1番目のグループ」に含まれる位置が存在しなくなった後に、「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つ目の位置の数字」まで繰り返す。

3.2 活性値計算とパラメータ

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、記憶内に含まれる全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る。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間が検索条件となる。そのため条件の数は2である。

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、両者の適合が存在しない場合には-1.0が設定される。両者に何らかの類似がある場合には、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

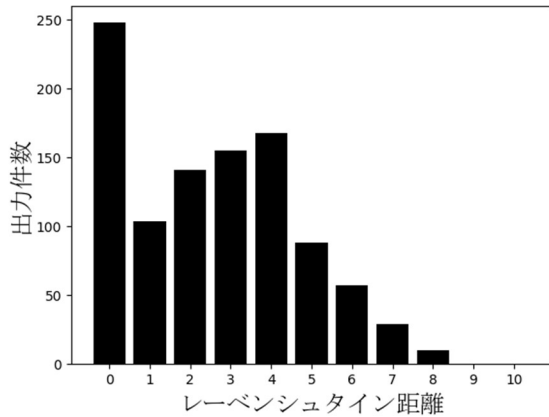


図2 パラメータのデフォルト値による結果

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、この値が小さな場合には、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 はロジスティック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が大きき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。

3.3 デフォルトシミュレーション

ACT-R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試行した。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、3.2で示したように、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。

4 データフィッティング

4.1 目的と方法

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個人のデータに対して、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。

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 (0.5~5.0, 0.5刻み)、オMISSIONエラーに関わる:rt (-6.0~6.0, 2.0刻み)である。これらの値をグリッドサーチし、各個人と最もよく対応するパラメータの組み合わせを探索した。

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70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。さらに、各参加者が報告した30の回答に対して、図1と同様に課題となった数列とのレーベンシュタイン距離のヒストグラムを構成した。そして、各参加者のヒストグラムに対して、もっと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。ヒストグラムインターセクションの値が最大となったモデルパラメータをその個人の内部状態を表すパラメータとして出力した。

4.2 結果

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

表 1. 個人モデルとアンケート回答の相関

	HI	mp	rt
年齢	0.006	-0.060	0.041
性別	0.015	-0.048	-0.066
体調	-0.329*	-0.318*	0.009
最終学歴	0.285*	0.074	0.107
暗記力自己評価	0.234	0.030	0.096
強気な	-0.030	-0.099	0.106
やる気がわいた	-0.031	-0.043	0.079
活気のある	-0.106	0.107	-0.025
熱狂した	0.040	-0.275	0.098
興味のある	0.032	-0.127	-0.023
興奮した	-0.030	-0.375*	0.082
誇らしい	-0.022	-0.213	0.321*
機敏な	0.097	0.082	-0.059
決心した	0.037	-0.262	0.029
注意深い	0.084	0.145	0.079
ポジティブ	-0.005	-0.169	0.125
恐れた	-0.029	-0.363*	-0.012
おびえた	-0.072	-0.340*	0.034
うろたえた	-0.121	-0.303*	-0.152
恥ずかしい	0.051	-0.212	-0.025
うしろめたい	-0.091	-0.301*	-0.191
びりびりした	-0.154	-0.223	-0.106
苦悩した	0.009	-0.196	-0.151
イライラした	0.129	-0.033	-0.204
神経質な	-0.011	-0.043	-0.125
敵意を持った	0.018	-0.312*	0.001
ネガティブ	-0.030	-0.274	-0.138

*は無相関検定の結果 $p < .05$

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 70 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。

表をみれば、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

情評価項目を見つけることができる。HI に関しては最終学歴の間で正の相関、体調との間で負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、体調を悪く報告した参加者、最終学歴が高い参加者において、適合するモデルが得られたと考えられる。

体調に関しては、:mp との間でも負の相関が観察される。この結果から体調を良いと報告した参加者ほど、コミッションエラーを起こしやすいモデルとよく適合したことになる。一般に何らかの認知的課題の成績は、体調の良さと正の相関を有すると考えられるため、この結果に関しては直観と反するものといえる。しかし、この結果に関しては、体調報告のタイミングが影響している可能性がある。これらのアンケートは実験終了後に行った。そのため、コミッションエラーを少なく課題を終えた参加者ほど、課題後の体調を悪く報告する傾向にあった可能性が考えられる。

また、:mp は、1 つのポジティブな感情評価項目（興奮した）、5 つのネガティブな感情評価項目（恐れた、おびえた、うろたえた、うしろめたい、敵意を持った）との間で負の相関を示した。これらの感情評価が高い人ほど、検索条件と適合しないチャンクに対するペナルティが小さく、コミッションエラーを起こしやすかったと解釈できる。この結果に関しても、実験後に興奮、もしくはネガティブな感情を抱いた参加者ほどコミッションエラーを起こしやすかったと解釈できる。

また、オMISSIONエラーと関連する:rt に関しては誇らしいと正の相関がみられた。すなわち、実験後に高い誇らしさを持った人ほど、検索される記憶の活性値に対する閾値を高く設定したと解釈される。高い閾値の設定は、活性値の低い記憶を検索することに対して労力をかけない個人と対応するものである。そのため、今回の実験において、実験後に誇らしさを保った参加者は、実験の参加に対して大きい労力を払わなかった可能性がある。

5 まとめ

本研究では記憶に関わるエラーの表出から、エラーに関わ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を実施し、その実験結果に対する組織的なパラメータフィッティングを個人のデータに対して実施した。その結果、実験後に報告された複数の個人属性や感情評価とモデルパラメータとの間で有意な相関が観察された。この結果より、本研究の手法により、表出された記憶のエラーからの内部状

態の推定が一定程度可能であることが示唆された。

しかし、本研究には実験で収集したデータ数、パラメータ範囲の設定に関する根拠の不足、データフィッティングにおける精度の不足など複数の限界がある。これらの問題を改善することで、将来的にはより頑健な個人のモデル化が可能であると考ええる。そして、そのような個人化されたモデルを利用することで、各個人の認知的な特性を踏まえたエージェントの設計が可能になると考える。

参考文献

- [1] Schacter, D. L.: The seven sins of memory: How the mind forgets and remembers, Houghton Mifflin harcourt, (2001)
- [2] Anderson, J. R.: How can the human mind occur in the physical universe? Oxford university press, (2007)
- [3] 川人潤子, 大塚泰正, 甲斐田幸佐, 中田光紀: 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, (2012)
- [4] Bothell, D.: Unit 5: Activation and Context, ACT-R tutorial (Version 7.27.7), (2022)