

認知アーキテクチャを組み入れた写真スライドショーの開発： 展望と課題

Development of a Photo Slideshow Implementing a Cognitive Architecture:
Visions and Challenges

森田純哉*¹
Junya Morita

平山高嗣*¹
Takatsugu Hirayama

間瀬健二*¹
Kenji Mase

山田和範*²
Kazunori Yamada

*¹名古屋大学
Nagoya University

*²パナソニック株式会社
Panasonic Corp

We are developing a photo slideshow app that can be used in life review and reminiscence. This app is characterized as an application of cognitive architecture, ACT-R (Adaptive Control of Thought-Rational [1]). Especially, in our study, the declarative memory of the architecture is developed from a personal photo library to make the app simulate the user's associative memory retrieval process concerning photos. We have so far conducted preliminary simulations to explore relations between ACT-R parameters and model behaviors. In this report, we will discuss a methodology of monitoring and support to enhance the elderly's health and longevity based on cognitive modeling approach.

1. はじめに

少子高齢化の進行する我が国の未来を、持続可能で活力あるものとするためには、健康寿命の延伸が必須である。我々は、健康寿命の延伸に寄与する心身状態を、「健康長寿力」と呼び、そのモニタリングとケアに役立つ情報システムの開発に取り組んでいる。我々の考える健康長寿力は、知力・気力・体力の3つの「力」により構成される。知力は認知機能（大脳皮質）、気力は情動機能（辺縁系・基底核）、体力は身体機能（自律神経系や循環器系）に対応する。近年の神経科学は、これらが密接に関係し合うことを指摘する [2]。よって、この3つを調和した形で保つことが、健やかな生活の持続に必要と考える。

本稿において、著者らは、上記の長期的目標に対し、認知アーキテクチャを利用するアプローチを示す。ここでいう認知アーキテクチャは、認知モデリングのベースとなるソフトウェアであり、脳機能と対応するモジュール群と、各モジュールの挙動に影響するパラメータ群から構成される。様々な認知アーキテクチャが提案されているものの、本研究ではベーシックなアーキテクチャとして、ACT-R[1]を用いる。

ACT-Rは、健康長寿力を構成する力のうち、知力のモデル化に適している。過去、ACT-Rは、認知心理実験をシミュレートする数多くの研究で用いられてきた。また、生理学的にいえば、ACT-Rのモジュール構造は、大脳皮質と基底核のループに対応し、強化学習的な意味での気力をモデル化する。さらに、近年、ACT-Rと自律神経系のモデルを接合する研究がなされ [3]、このアーキテクチャは、身体と認知の関係のモデル化をも可能にしている。

ACT-Rによるモデリングの対象課題として、回想法などで行われている自伝的記憶の振り返りに焦点を当てる。自伝的記憶の振り返りは、関連する記憶を連続的によびおこすメンタルタイムトラベルと形容される意識状態を引き起す [7]。また、輝かしい過去の追憶により、ポジティブな感情が呼び起こされるとも言われる [5]。

最近のデジタル環境では、ユーザの自伝的記憶が、ライフログとして記録される。ライフログのなかでも写真は、記憶の振り返りに有効である。こういった背景から、著者らは、ACT-R

を組み入れた写真スライドショーを開発している。ACT-Rを用いることで、ユーザとなる高齢者のパーソナルな思い出のモデルを構築する。そのモデルによって、写真に対する高齢者の反応を予測しつつ、記憶を活性化させ、気力を充実させ、活動的な社会生活へ誘導する写真を提示していく。

図1は、実装中の写真スライドショーの構成を示している。写真はコンシューマ向け写真管理ソフトに格納される。後述するプロトタイプでは、iPhoto 9.5が用いられる。iPhotoのメタデータ、および写真に対する画像処理によって、写真の情報を記号化したデータベースを構築する。写真の表示は、Webサーバによって制御される。WebサーバがACT-Rのプロセスのトリガーを引き、ACT-Rからのレスポンスを得ることで、写真をブラウザに連続提示する。

この状況は、ユーザがブラウザの写真を観察すると同時に、ACT-Rが同一の写真を観察し、次の写真を検索する状況に対応する。ここでACT-Rのプロセスがユーザのプロセスをトレースできていれば、ACT-Rによって検索される写真は、同調メカニズムによって、ユーザにポジティブな効果を与えると期待できる。さらに、ユーザに同調しつつも、高齢者が自発的には思いつきづらい写真を提示することで、知力や気力を望ましい状態へ導くことも可能になる。

こういった仕組みを実現するには、(1)ユーザの記憶をトレース可能なモデルを構築する、(2)ユーザからのフィードバックによってモデルを調整するという2つの課題がある。本稿では、前者の課題に焦点をあてる。すなわち、ACT-Rによる自伝的記憶の検索の性質を探り、その可能性と課題を明らかにすることを旨とする。

2. ACT-R モデル

本研究では、ACT-Rの視覚モジュール、宣言的モジュール、ゴールモジュール、プロダクションモジュールを用いることで、写真を知覚し、関連する写真を連続して連想検索するモデルを構築した。

2.1 視覚モジュール

ACT-Rの視覚モジュールは、仮想的なディスプレイ (AGI: ACT-R Graphical Interface) と相互作用する。ただし、AGIに置かれる要素 (ボタン、テキストなど) は、事前に記号

連絡先: 森田純哉, 名古屋大学未来社会創造機構, 名古屋市千種区不老町, j-morita@cmc.ss.is.nagoya-u.ac.jp

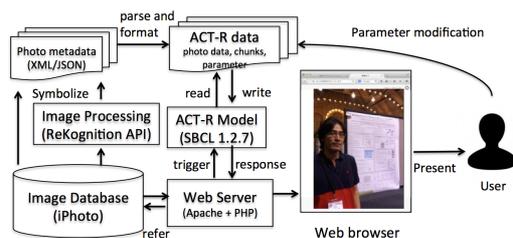


図 1: 写真スライドショーの構成

- (face753 isa meaning text "face753")
- (s-broom isa meaning text "broom")
- (<GUID> isa meaning text "<GUID>")
- (*** isa include pdata <GUID> people face735)
- (*** isa include pdata <GUID> scene s-collage)
- (*** isa time pdata <GUID> time time53)
- (*** isa geo pdata <GUID> place geo30)

図 2: チャンクの例

化される必要がある。本研究では、ACT-R への入力として、既存の画像処理エンジンの出力を用いる。具体的には、iPhoto に付属する顔認識の機能、および ReKognition API (<https://rekognition.com>) によるシーン認識を利用する。ReKognition API は、Deep Learning ベースの認識エンジンとされ、入力された写真に対して、山、海、食べ物などの言語的なタグを出力する。

上記の画像認識によって抽出される人とシーンの情報を、AGI の 2 次元空間に配置する。例えば、図 1 の写真は以下のようにコーディングされる。

```
(setf *photo-struct*
  '(("QM7QZbQMSNSxq4oS7CbS1A" 10 10 BLUE)
    ("text" 0 430 GREEN)
    ("cloth" 128 430 GREEN)
    ("face" 256 430 GREEN)
    ("girl" 384 430 GREEN)
    ("crosswordpuzzle" 512 430 GREEN)
    ("face4701" 91 245 RED)))
```

この例で、*photo-struct* は 7 つのリストから構成される。各リストの第一要素が AGI に配置されるテキスト、第 2・第 3 要素はテキストを配置する AGI の座標、第 4 要素はテキストの色を示している。色はタグの種類に応じている。BLUE は写真 ID、GREEN は ReKognition API によるシーntag、RED は iPhoto データベースにおける人物 ID を示している。写真 ID は、写真にシーンや人以外の固有情報が含まれると考え、AGI に並べる。

2.2 宣言的モジュール

ACT-R の宣言的記憶を構成する要素はチャンクと呼ばれる。図 2 にモデルのもつチャンクの例を示す。

図 2a-c は、AGI に置かれるテキストの意味を記述する。これらを利用することで、モデルは写真の情報を認識する。図 2d-g は、写真の属性を示す。写真の属性として、What 属性、Who 属性、Where 属性、When 属性を想定する。写真に含まれる顔 ID を Who 属性 (図 2d) が記述し、シーntag を What 属性 (図 2e) が記述する。When 属性 (図 2f) と Where 属性 (図 2g) は、それぞれ写真の撮影された時期と場所を示す。これらの属性は、Exif に含まれる数値データのクラスタリングによってコーディングされる。

2.3 ゴールモジュール

ACT-R のゴールモジュールは、課題の状態を一時的に保持するバッファである。本研究のゴールモジュールは、モデルの状態を表すスロット (State)、現在表示されている写真の ID を格納するスロット (PhotoID)、次に表示する写真 ID を格納するスロット (NextPhoto)、現在表示されている写真の属性を格納するスロットにより構成される。このうち、What 属性と Who 属性を格納するスロットは、それぞれ容量 3 のプッシュダウンスタックとする。これらスロットの値は、以下に示すプロダクションモジュールによって逐次的に埋められる。

2.4 プロダクションモジュール

モデルは、ブラウザに表示されている写真と関連する写真を宣言的記憶から検索し、検索された写真をブラウザに表示する。つまり、自身の検索した写真が次の手がかりとなり、写真検索を繰り返していく。このプロセスは、イベントに関する記憶を自由連想していくことに対応する。

図 3 にモデルの持つルールを示す。ルールは矢印で結合された複数の独立したプロセスを構成する。各プロセスが終了した時点で、ゴールバッファの State スロットが start に切り替えられ、新たなプロセスが選択、開始される。プロセスの選択は、先頭のルールの IF 節に記述される条件によって制御される (図中では、When... と表示)。条件が共通するプロセスが複数あった場合、競合解消がなされる。ACT-R の競合解消は、ルールに付与されたユーティリティの比較による。本研究では、全てのルールに対して同じユーティリティの値を設定した (ランダムなプロセスの選択)。

図 3 上段のボックスは、写真から顔、シーン、写真 ID を認識するプロセス群である。start-perceive ルールにより、AGI の文字列がランダムに選択される。選択された文字列に対応するチャンク (図 2a-c) が宣言的モジュールから検索され (perceive-photoID, Perceive-scene, perceive-face)、その値をゴールモジュールに格納する (photoID-to-goal, scene-to-goal, face-to-goal)。写真に直接現れない時間と場所は、photoID がゴールバッファに格納された後に、認識できる (図 2 中段のボックス)。下段のボックスは現在の写真と関連する写真を思い出すプロセスである。現在の写真と共通する属性を持つチャンク (図 2d-g) が検索され、ゴールモジュールの NextPhoto が埋められる。

モデルは、上記のプロセスの選択と実行を、一定時間のトライアルのなかで繰り返す。トライアル中に、上段や中段のプロセスが選択されるごとに、ゴールのスロットが埋められ、写真に対する認識が深まっていく。よって、トライアルの持続時間が長ければ長いほど、あるいはモデルの処理が早ければ早いほど、より豊かな情報に基づく写真検索が行われる。

2.5 サブシンボリック計算

宣言的記憶の検索では、プロダクションモジュールによる検索リクエストに合致した全てのチャンクに対して活性値が計算され、最も活性値の高いチャンクが検索される。活性値 A_i は、ベースレベル活性値 B_i 、活性化拡散 S_i 、ノイズ ϵ_i の合計として計算される。

$$A_i = B_i + S_i + \epsilon_i \quad (1)$$

ベースレベル活性値の計算において、本研究は以下の式を用いる。

$$B_i = \ln(n/(1-d)) - d * \ln(L) + \beta_i \quad (2)$$

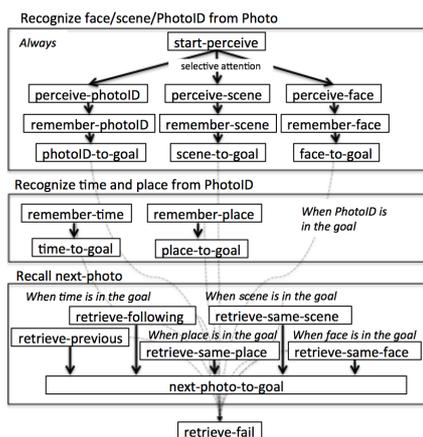


図 3: モデルのプロセス

n はチャンク i の出現回数, L はチャンクが作られてからの時間, d は減衰率, β_i はオフセット値を示す.

n を含む第一項は学習曲線を描き, L を含む第二項は忘却曲線を描く. ベースレベル活性化値は, 2つの対数曲線の差分 (オッズ差) である. 本研究では, L を写真の撮影時刻によって設定した. AGI のテキストとシーンや顔の関係を記述するチャンクは, そのタグの出現回数に応じた値を n に割り当て, 写真 ID を含むチャンクは, n を 1 とした. つまり, 写真撮影後, 一度もその写真を観察していない状況をシミュレーションの条件とした.

式 1 の第二項, 活性化拡散は現在のコンテキスト C に対するチャンク i の連想強度 S_i として計算される.

$$S_i = \sum_{j \in C} W_j S_{ji} \quad (3)$$

C は, ゴールバッファに含まれる属性値 j の集合を表す. W_j は, 属性値 j に付与される注意の重みを表し, S_{ji} は, 属性値 j と宣言的記憶のチャンク i の連想強度を表す.

図 4 に, 本研究における W_j と S_{ji} の計算を説明している. 赤色の矢印は, When 属性 $t2$ を使った検索リクエストを示している. ゴールモジュールには, 検索リクエストに利用されない属性値 ($g2, e, d$) も格納されており, これらもコンテキスト C に含まれる. 重み W_j は, パラメータ GA (Goal Activation) をコンテキストの要素数で割ることで定まる.

図 4 の下段は, 宣言的記憶の状態を示す. 検索リクエストに対し, When 属性 $t2$ と写真 ID3 を結合するチャンク i がマッチしている (図 2f の形式のチャンク. 図 4 ではオレンジのセルとして表示). さらに, 本研究における活性化拡散の計算では, 写真 ID3 を項とする宣言的記憶の全チャンクを結合し, 仮想チャンク i' を生成する (赤枠のセル). 仮想チャンク i' に, コンテキスト要素 j が含まれれば, 両者に連想リンクが貼られ, リンク強度 S_{ji} が計算される.

$$S_{ji} = MAS - \ln(fan_j) \quad (4)$$

MAS (Maximum Associative Strength) は, 連想強度の最大値であり, fan は, 要素 j と連想関係にあるチャンクの宣言的記憶における総数を示す. MAS から fan を引くことで, 宣言的記憶の中で多くのチャンクと結合する中心性の高い要素

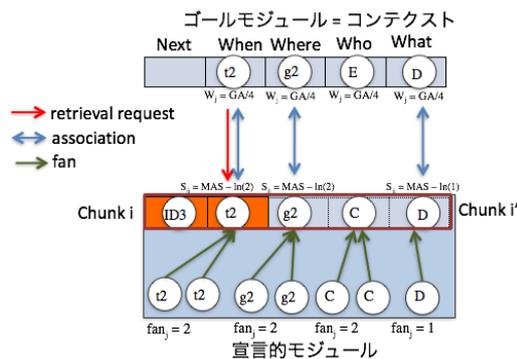


図 4: 活性化拡散

の連想強度が低くなる. つまり, 活性化拡散の計算は, 検索にコンテキストの効果を含めるだけでなく, 宣言的記憶のネットワークにおいて, 中心性の高い要素への検索の集中を防ぐ効果も有している (fan 効果).

3. シミュレーション実験

3.1 シミュレーション設定

3.1.1 データセット

第一著者が私的に保有する写真 3202 枚を利用した. これらは, 本研究のために選択したものではなく, 私生活の中で撮影した, あるいは第一著者が参加したイベントにおいて他者が撮影し, 譲渡を受けるなどしたものである. 2015 年 1 月 1 日時点で利用可能な全ての写真を対象とした. 撮影時にデジタル化されたものだけでなく, フィルム写真についてもスキャナで取り入れ, 手作業で場所と日付の情報を入れてある.

写真は, iPhoto で管理されており, 検出された全ての知人の名前が入力され, 全ての写真について, ReKognition API によるシーン認識の結果を得ている. 日付は, 1977 年から 2014 年 12 月 13 日までの範囲である.

3.1.2 手続き

ACT-R の組み込み関数を用いることで, モデルのシミュレーション時間を, 2015 年 1 月 1 日 00:00:00 に設定した. この設定とチャンクに付与されたパラメータ L の組み合わせにより, 現実のタイムスケールと対応した記憶検索を行う.

一枚の写真が提示されるトライアルを 5 秒間とし, モデルは写真 200 枚を連続して検索した. これを一つの実行とし, 20 回繰り返した. いずれの実行においても, iPhoto ライブラリの最新の写真を最初にモデルに与えた.

3.1.3 パラメータ設定

はじめに, ACT-R のデフォルトのパラメータ設定によって, モデルを走らせた. 結果, モデルは全てのトライアルにおいて, 写真の検索に失敗した. この問題に対し, 本研究は全てのチャンクのベースレベルが正の値となるよう, Base Level Constant (BLC; 式 2 の β) を設定することで対処した ($BLC = 15$).

BLC を固定した上で, 式 1 の ϵ (ANS: Activation Noise S) と式 4 の MAS を操作した. 2 水準の ANS 条件 (the low noise condition: $ANS = 0.2$, the high noise condition: $ANS = 1.0$), 2 水準の MAS 条件 (the no spread condition: $MAS = 0$, the high spread condition: $MAS = 10$) を設定した. 2 つを組み合わせた 4 つの条件において, モデルによる検索の特性を検討した.

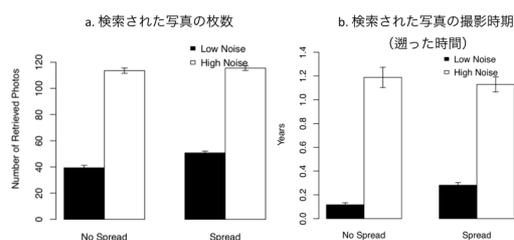


図 5: シミュレーションの結果

3.2 結果

結果を図5にまとめた。図5aは各条件において検索された写真の異なり数、図5bは、各条件において平均して何年前の写真が検索されたかを示している（いずれも20回の実行の平均）。図より、活性化値の計算がベースレベルに大きく依存する条件（Low Noise/No Spread）で、最近の同じ写真が繰り返し検索されたことがわかる。ベースレベル活性化値に加え、活性化拡散とノイズを活性化値に付与すると、より多様で古い記憶の検索がなされるようになる。特に、ノイズの影響は、図5a、図5bの両者で顕著に現れている。それに対して、活性化拡散の効果は高ノイズ条件において観察されず、低ノイズ条件においてのみ観察された。

3.3 考察

シミュレーションが顕著に示すのは、ベースレベルの計算によって引き起こされる直近の特定の記憶への束縛である。ベースレベル活性化値の計算（式2）は、最近の記憶を優先的に検索させ、一度検索されたチャンクの活性化値を強化する。こういったACT-Rの特性は過去の研究でも指摘されており、“*pathological behaviors such as out-of-control looping*”などと形容されている[4]。

ただし、このような病的な振る舞いに心理学的な妥当性がないとも言えない。Schacterは、記憶のつきまといを人間の記憶のエラーの一つに挙げている[6]。また、認知症など加齢による疾患において、同じ話が繰り返されるなどの症状も頻繁に観察される。よって、ACT-Rによる病的な検索のループが、特定のユーザの状態をモデル化する可能性はある。

また、本研究のシミュレーションは、こういった病的なループから脱出する方法として、活性化拡散とノイズを付与する方法を示した。このうち、ノイズの影響は顕著に現れた。ノイズの付与により、ベースレベルによる束縛が薄れ、より多くの古い記憶が検索された。

ここで生じる疑問は、ノイズの意味である。認知アーキテクチャに関する近年の研究では、アーキテクチャのパラメータを、情動の観点から説明する提案がなされている。それらの提案で共通する考えは、情動がパラメータの調整器となっているということである。特に、宣言的記憶のノイズについては、覚醒度やストレスと対応づける議論がなされている[3]。こういったノイズに関する議論から導かれる本研究の示唆は、古い記憶の追憶は、覚醒度の高いとき、集中しているときには起きないという予測である。

しかし、本研究の結果は、あまりに高いノイズの付与は、健全な認知プロセスを妨害することも示唆する。図5において、活性化拡散の効果は、高いノイズの付与によって消滅した。活性化拡散は、コンテキストと関連した記憶の想起を助けつつ、中心性の高い記憶への過度の集中を防ぐ利点を持つと考えら

れる。高いノイズの付与によってこの効果が消滅するのであれば、ノイズの範囲を適切に調整するメカニズムも必要になる。

4. 結論と課題

本研究の最終的な目標は、認知モデルによるモニタリングとケアを組み入れた高齢者支援システムを構築することである。この目標に向け、本研究ではACT-Rによる写真検索を検討するシミュレーションを行った。

本研究の成果に対し、未だ多くの課題が残される。まず、扱うデータの充実が不可欠である。今回利用したデータセットは第一著者のものに限られていた。本研究の背景を踏まえれば、より長い人生の記録を持つ高齢者のデータを収集しなければならない。モデルの妥当性を検証するための心理学的な実験も必要になってくるだろう。

本研究のモデル構成に対しても、検討が必要な課題が多く残されている。特に、本研究では直接検討の対象としなかったパラメータ、BLC、減衰率、チャンクの利用回数 n については、その設定の妥当性を十分に検討する必要がある。心理学的・生理学的に裏付けられたパラメータセットと、モデルの多様な振る舞いを対応づけていくことで、健康長寿力をモニタリングするユーザモデルの構築が成し遂げられると考えている。

謝辞

本研究は独立行政法人科学技術振興機構（JST）の研究成果展開事業「センター・オブ・イノベーション（COI）プログラム」の支援によって行われた。

参考文献

- [1] J. R. Anderson. *How can the human mind occur in the physical universe?* Oxford University Press, New York, 2007.
- [2] AR Damasio. *Looking for Spinoza: Joy, sorrow, and the Feeling Brain*, 2003.
- [3] C. L. Dancy and F E Ritter. Using a cognitive architecture with a physiological substrate to represent effects of a psychological stressor on cognition. *Computational and Mathematical Organization Theory*, in-press.
- [4] C Lebiere and B. J Best. Balancing long-term reinforcement and short-term inhibition. In *Proceedings of st Annual Conference of the Cognitive Science Society*, 2009.
- [5] C Routledge, T Wildschut, C Sedikides, and J Juhl. Nostalgia as a Resource for Psychological Health and Well-Being. *Social and Personality Psychology Compass*, Vol. 7, No. 11, pp. 808–818, 2013.
- [6] D L Schacter. *The Seven Sins of Memory: How the Mind Forgets and Remembers*. 2002.
- [7] E Tulving. Memory and consciousness. *Canadian Psychology/Psychologie canadienne*, Vol. 26, No. 1, pp. 1–12, 1985.