

統合認知アーキテクチャの魅力と困難: ACT-R を利用した研究事例

Joys and sorrows of using integrated cognitive architectures:
Example studies of using ACT-R

森田 純哉

Morita Junya

北陸先端科学技術大学院大学知識科学研究科
Japan Advanced Institute of Science and Technology

Abstract: This article discusses the role of ACT-R in the current cognitive science. ACT-R is one of the unified theories of cognition, which integrates cognitive modules such as memory, perception, motor skills and goals. This architecture recently have included sociality, brain, physiology, and emotion in the scope. This article also presents two studies using ACT-R. Through presenting these studies, the benefits and the difficulties of this architecture will be discussed.

1. はじめに

統合認知アーキテクチャという言葉は、Newell によって用いられ始めた (Newell, 1994). 彼自身関わった Soar をはじめ、これまでに多数の統合認知アーキテクチャが構築されてきた (レビューとして Samsonovich, 2010). 本稿は、その中でも、John. R. Anderson らによって主導される ACT-R (Adaptive Control of Thought – Rational: Anderson, 2007) に焦点をあてる。

統合認知アーキテクチャに関する Anderson の考えは、2012 年 3 月現在 ACT-R の WEB サイト (<http://act-r.psy.cmu.edu/>) にて閲覧できる。2011 年度 Franklin medal award の受賞式において、彼は、細分化された認知心理学の現状に対するフラストレーションを率直に述べた。認知心理学は、心の機能を知覚、記憶、思考、学習などに分解することで、個別の研究を遂行してきた。そのようなアプローチは、個別機能の詳細な働きを厳密に検討するという点で利点がある。しかし、個々の研究領域間の繋がりは、希薄になりがちである。ACT-R は、細分化された認知機能の研究を統合し、人間の複雑な活動を説明しようとする。

本稿では、心の科学に関する近年の研究動向をながめ、ACT-R による心の科学の統合を示す。また、著者自身の研究を示し、ACT-R の魅力と

困難さを伝えることを試みる。

2. 心の科学の進展と ACT-R

2.1. 脳・認知・社会

認知科学をはじめとする心の諸科学は、今世紀に入り、急速な進展をみせた。安西 (2011) は、その進展を、脳、認知、社会という異なるレイヤを統合する動きと捉えている。認知機能が脳にマップされ、認知機能から構成される社会性の研究が盛んになった。その一貫として、感情に関する身体的・進化的な研究も盛んになされるようになってきた。

しばしば、脳や社会性と関連する研究の流れは、シンボリックな認知プロセスに関する研究 (古典的な認知科学) と対比的に語られる。ここで述べるシンボリックな認知プロセスとは、概念や外的な事物を記号的に表現し、心の流れを記号操作と捉えるアプローチである (Newell and Simon, 1976). 認知神経科学に関わる文献の中で、このようなシンボリックなアプローチは、たびたび否定的に扱われる。

Anderson 自身は、ACT-R をシンボリックなシステムと位置づけているわけではないようである。その著書の中で彼は、心のプロセスが、記号か非記号的かを争う論争は無意味であると述べている (Anderson, 2007). しかし、ACT-R は、一般的な数理・計算モデルと異なる構成を

持ち、知識や事物を記号的に表現する。果たして、このようなアプローチは、近年の心の科学において、どのような役割を果たすのであろうか。この問いに答える前に ACT-R の構成について簡単に説明をしておきたい。

2.2 ACT-R の構成

ACT-R はモジュールの集合として認知を説明する。中央に配置されるプロダクションシステムに、周辺的なモジュールが結合される構成である。周辺的なモジュールには、外界と相互作用するための知覚-運動モジュール、宣言的記憶を貯える宣言的モジュール、現在の課題の状況やゴールを表現するゴールモジュールが含まれる。各々のモジュールはバッファを介してプロダクションシステムと結合する。プロダクションシステムは、モジュールのバッファを条件として、プロダクション (IF-THEN ルール) を発火させ、その内容を書き換える。

ACT-R はさらに、複数の学習メカニズムを統合する。その学習メカニズムは、大きくシンボリック学習、サブシンボリック学習に分けられる。シンボリックな学習として、新たな宣言的知識を蓄積する事例ベース学習、宣言的知識の内容を手続き化するコンパイル学習が備わる。サブシンボリック学習として、プロダクションの発火を制御する効用値の更新 (強化学習)、宣言的知識の活性値の更新が含まれる。

各モジュールの動作 (プロダクションの実行時間や宣言的知識の検索時間)、学習に関わるパラメータ (学習率や忘却率) は、過去の心理実験の知見を反映することで推定されている (Anderson, Bothell, Byrne, Douglass, Lebiere, & Qin, 2004)。このように、ACT-R は、認知心理学の研究の成果をアーキテクチャのパラメータ設定として統合する。これにより、ACT-R は、たとえば実時間と対応するシミュレーション時間の見積もりなどを可能にしている。

2.2 脳・ACT-R・社会

本章の冒頭において、今世紀の心の科学が、脳、認知、社会を統合する方向に進んでいることを述べた。近年の ACT-R の発展もこの流れに従う。つまり、ACT-R は、現在、認知機能間の統合だけでなく、レイヤをまたぐ統合に挑戦をしている。

脳と認知の統合として、モジュールと脳部位

とのマッピングがなされている。たとえば、プロダクションシステムは大脳基底核に対応づけられ、宣言的モジュールは海馬に対応づけられる。宣言的モジュールに付属するバッファは、前頭前野に対応づけられる。このような対応づけを仮定することで、たとえば宣言的知識が手続き化されていくときの脳内プロセス (基底核へ活動が推移していくプロセス) が説明される。最新の研究では、脳部位のニューラルネットモデルと ACT-R の等価性を示すこともなされている (Stocco, Lebiere, O'Reilly, and Anderson, 2010)。

さらに、人間の身体に関する生理学的なモデルと ACT-R を統合する試みもなされている。Dancy, Ritter, and Berry (in press) による ACT-R の拡張では、循環器系や体温などの身体的状態に関わる数理モデルが ACT-R に統合される。身体的なモデルは、宣言的知識やプロダクションの選択、あるいは学習に関わる ACT-R のパラメータを調整する役割を果たす。このような仕組みにより、Damasio (1994) が述べたような情動と意思決定の関係についても説明の射程に収まる。

認知と外界との結びつきに関する研究も盛んになされている。その1つの潮流は、エルゴノミクス的な研究である。たとえば、Salvucci はドラビングシミュレータを操作するモデルを構築し、ドライビング時のマルチタスクの影響をシミュレーションしている (Salvucci, 2006; Salvucci, & Taatgen, 2010)。

認知と外界の結びつきに関する別の潮流が社会性に関する研究である。近年、ACT-R を用いた社会シミュレーションが盛んに行われるようになってきた。Reitter and Lebiere (2011a) は、新たな言語が共同体の中で構成され、普及するプロセスに関するモデルを構築した。彼らの研究は、ACT-R による社会シミュレーションをネットワークサイエンスに統合することを目指している (Reitter & Lebiere, 2011b)。

上記の一連の研究に示されるように、シンボリックな表現を用いる ACT-R は、心の科学の最先端の問題に適用するツールとして今なお力を発揮している。以下では、このことをより具体的に示すために、著者らが現在行っている研究を2例示す。個別の研究の詳細は、Morita, Miwa, Maehigashi, Terai, Kojima, and Ritter (2011)、および Morita, Konno, and Hashimoto (submitted)

を参照されたい。

3. 研究事例

3.1 知覚と運動を介した自動化システム利用の意思決定モデル

3.1.1 背景と目的

自動化システムは、人間の作業を機械が肩代わりする。現代の航空機や船舶の運行において、自動的な操縦は一般的なものとなっている。また、我々にとってより身近な自動車の操縦においても、自動的な運転システムの開発がなされている。しかし、自動化システムの利用に関わる判断は、時としてクリティカルな結果をまねくことが指摘されている。

自動化システムの利用に関わる意思決定についての数理的なモデルが構築されている。

EDFT (Extended Decision Field Theory: Gao & Lee, 2006) は、自動化システムに対する信用と手動操作に対する自信のバランスによって、自動化システムへの依存を説明する。信用と自信の程度は、自動化システム、手動操作のそれぞれの性能に応じて動的に変化する。

EDFT モデルの限界は、外界とのインタラクションをモデルの対象に含めないことである。このモデルにおいて、性能は所与のパラメータとしてモデルに与えられる。よって、このモデルを外界との相互作用に応じてシステム性能が変動する状況に適用することはできない。そこで、本研究では、ACT-R を用いることで、外界との相互作用を含む状況に EDFT を拡張することを目指した。

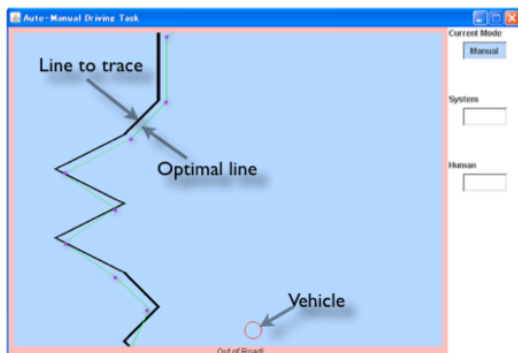


図 2: 路線追従課題

3.1.2 タスク

Maehigashi, Miwa, Terai, Kojima, & Morita, (2011) による路線追従課題を対象にモデルを構築した。この課題では、スクロールする路線

をヴィークルでトラッキングすることを実験参加者に求める (図 2)。その際、オートモードとマニュアルモードという 2 つの操作方法が参加者に提供される。オートモードは、自動操作プログラムがラインをトレースする。マニュアルモードは矢印キーの操作によって参加者自身がヴィークルを操作する。課題中、実験参加者は 2 つのモードをスペースバーを押すことで自由に切り替えることができる。

Maehigashi らの実験では、いずれのモードにおいても、トレース性能が操作された。低いトレース性能のとき、ヴィークルの動きは鈍くなり、路線の追従が困難になった。実験参加者は、明示的に 2 つのモードの性能を知らされなかった。2 つのモードを主体的に切り替えつつ、その場の状況において有利なモードを探ることを迫られた。

3.1.3 モデル

EDFT モデルにおける信用と自信は、プロダクションの効用値と対応するものとみなせる。つまり、マニュアルモードの利用に関わるプロダクションの効用値を自信、オートモード選択時に発火するプロダクションの効用値を信用と考える。そして、2 つのモードの利用に関わるプロダクションが競合するモデルを構築する。ACT-R におけるプロダクションの効用値は、報酬により更新される。そのため、EDFT と同様に自動化システムへの依存の時間的変化をシミュレーションできる。

ただし、本研究では、EDFT モデルと異なり、課題とインタラクションするモデルを ACT-R によって構築した。構築されたモデルは、マニュアルモードにおいて、実験参加者と同じように、ヴィークルを操作する。また、プロダクション間の競合解消に応じて、スペースバーを押し下げ、モードを切り替える。

3.1.4 シミュレーション結果

オート・マニュアルモードのそれぞれに対して、5 レベルの性能水準を設け、シミュレーションを行った。その結果、モデルは、実験において得られた路線追従課題のパフォーマンス、課題中のオート使用率、モードの切り替え回数などのデータを良くシミュレートした。

3.1.5 示唆

シミュレーションの結果は、ACT-R により構築されたモデルが、人間の合理的な意思決定を再現できることを示す。ただし、この結果をも

って、ACT-R の有用性を主張することはできない。本研究において構築されたモデルは、EDFT モデルと同様の意思決定の構造を持っている。違いは、ACT-R によるモデルが知覚-運動モジュールを持ち、人間と同様のタスクを遂行することである。つまり、この研究は、数理的な意思決定のモデルと知覚-運動的なモデルを統合するものである。このような統合的アプローチにより、将来的には、マルチタスク状況下での自動化システムへの依存の問題を検討する予定である。

3.2 コミュニケーションのモデル

3.2.1 背景と目的

共有する言語がない状態でのコミュニケーションの成立過程を扱った。コミュニケーションの成立は、言語、あるいは社会性の起源に繋がる問題である。本研究では、コミュニケーションの成立に関与する要因として模倣に注目した。模倣による乳幼児の言語獲得について、これまでに多くの研究者が議論してきた (例えば Tomasello, 1999)。しかし、新たなコミュニケーションの成立に果たす模倣の役割について未だ十分な検討はなされていない。

3.2.2 タスク

Konno, Morita, and Hashimoto (in press) によって検討されたコーディネーションゲームを課題とした。このゲームは、独立したラウンドの繰り返しから構成された。ラウンドの流れを、図 3 に示す。

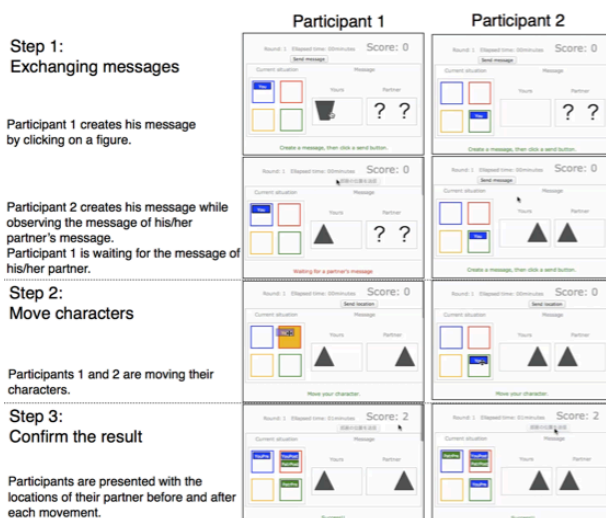


図 3: コーディネーションゲームの流れ

各ラウンドにおいて、2 名の参加者は、それ

ぞれ別の初期位置に配置されるキャラクターを共通の位置に移動させることを目指した。位置の移動は上下左右のみ許され、それぞれの参加者には、初期位置から移動することのできない位置が存在した。そのため、参加者は移動に先立ち、共通の移動先を決定する必要があった。参加者は、意味や運用ルールが事前に定められない図形 (■や▲) を横に 2 つ組み合わせたメッセージをパートナーへ送信できた。メッセージの交換は各ラウンドにおいて一度のみ許された。共通の位置への移動が成功した場合、参加者には得点が与えられ、次のラウンドに進んだ。一定の得点に至った段階で課題を終了した。

3.2.3 モデル

本研究におけるコミュニケーションのモデルは、2 体の ACT-R エージェントが、シミュレーション課題環境を介して相互作用するものである (図 4)。

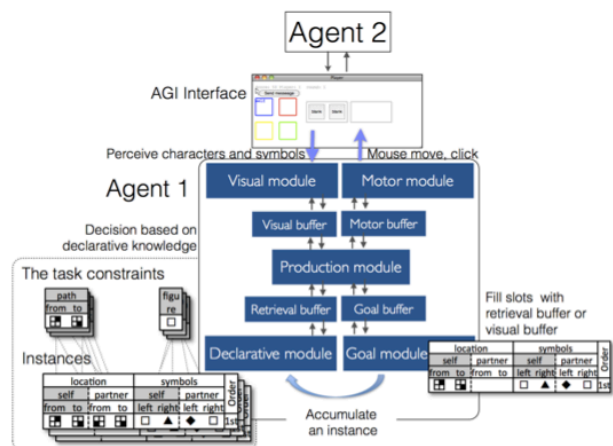


図 4: モデルの構成

エージェントの視覚モジュールは、キャラクターの位置やメッセージの内容を課題環境から取得する。また、運動モジュールは、課題環境に含まれるマウスを操作することで、メッセージを作成/送信し、キャラクターを移動する。ゴールバッファは、自分とパートナーの初期位置と移動先、およびメッセージを構成する左右の図形を格納するスロットをもつ。エージェントは、各ラウンドにおいて、知覚モジュール、あるいは宣言的モジュールから取得された値を用い、これらのスロットを埋めていく。各ラウンドの終了時に、スロットの値は全てリセットされる。

宣言的モジュールには、過去のラウンドにおけるゴールバッファの状態が、事例として格納

される。これらの事例は、移動先や送信メッセージの決定に用いられる。事例の適用は、直接的、あるいは模倣を介してなされる。いずれの場合においても、現在のゴールバッファの状態を手がかりとして事例が検索され、検索された事例によって、ゴールバッファの空白のスロット（移動先、メッセージを構成する図形）が埋められる。事例を直接的に適用する場合、現在のゴールバッファと合致する事例が宣言的モジュールから検索され、直接的にスロットが埋められる。模倣によって事例を適用する場合、自分とパートナーの位置が入れ替わった事例が検索され、位置を入れ替えた状態でスロットの値が埋められる。

決定に利用される事例は、成功したラウンド（エージェントの移動先が一致したラウンド）の終了時に獲得される（事例ベース学習）。また、事例に含まれる情報は、コンパイルを経ることで、プロダクションとして手続き化される。コンパイルされたプロダクションの効用値は、ラウンド終了時に付与される報酬により変化する。その結果として、コミュニケーションの成功に寄与する行動が強化される。

3.2.4 シミュレーション結果

直接的な事例適用のみを行うモデルと模倣を加えたモデルを比較した。結果、模倣が加わることで、モデルは素早くより精度の高い（正確に移動先を一致させる）コミュニケーションを成立させた。また、模倣を加えたモデルは実験によって得られた結果をより良く再現した。さらに、模倣によって、コミュニケーションに用いられるメッセージがエージェント間で同型に近づくことが示された。

3.2.5 示唆

シミュレーションの結果は、模倣がコミュニケーションの成立に大きな役割を果たすことを示す。学習された事例は、直接的な方法に加え、自己と他者を入れ替えた方法で利用できるようになる。これによって、事例の検索の成功率が模倣により上昇する。また、模倣を介することで、コミュニケーションに従事する個人の間で共通の「ことば」が構成されていく。

本研究において実装された模倣は、自己を他者の立場におく視点取得の一種である。ACT-R のアーキテクチャの上で、この視点取得のプロセスは、コンパイルによって自動化されていく。つまり、当初は宣言的モジュールを利用して意

識的になされる視点取得が、プロダクションシステムのみで自動的に行われるようになる。モジュールと脳部位とのマッピングを用いれば、海馬、前頭前野の利用から、大脳基底核のみの処理に推移していくことが推測される。現在、著者の所属するグループでは、この課題遂行中の脳波の解析が進められている。今後、解析された脳波データを対象に、脳内プロセスのシミュレーションをすることを予定している。

4. まとめ

4.1 ACT-R の魅力

本稿で示した 2 つの研究は、共に ACT-R を外界のシステム（自動化システム、他者）とつなぐものである。また、3.2 の最後で述べたように、著者は、脳内プロセスのシミュレーションを将来的に実施することを検討している。このように、ACT-R は認知の内と外を統合的に説明するポテンシャルを持っている。この統合性こそ ACT-R の第一の魅力である。ACT-R を用いることで、研究者は、心の科学の全体を俯瞰する地図を手に入れる。

また、3.1 に示したように、ACT-R は、具体的な人間の行動を再現する利点をもっている。ACT-R の知覚-運動モジュールは、人間の被験者が用いるのと同型の課題環境を操作し、モデルは、実時間と対応するシミュレーション時間を出力する。この利点により、ACT-R は、シミュレーションと実験との対応を明確にする。

ACT-R の用いるシンボリックな表現も魅力の 1 つである。先述したように、Anderson 自身はこの点に対して曖昧な立場をとっている。しかし、著者は、ACT-R のバッファに格納される記号表現は、我々の素朴に考える意識状態に近いものと考えている。ACT-R は、プロダクションシステムにより、意識状態の移ろいを明瞭に記述する。記号的な意識状態の内容、およびその遷移こそ、多くの心理学者や認知科学者が元來說明したかったものではないだろうか。

4.2 ACT-R の困難さ

上記の魅力は、同時に ACT-R を用いたモデルの困難さにも繋がる。人間の心は複雑であり、それを統合的、具体的に記述することは、説明やコーディングの肥大化を招く。一般的にプロダクションシステムは、独立した IF-THEN ルー

ルの組み合わせにより構成される。多くのルールをつなぎ、複雑なプロセスを形成することは困難な作業となる。この問題に関して、ACT-Rのコーディングを支援する研究も盛んになされている (e.g., Ritter, 2009; Zhao, Paik, Morgan, & Ritter, 2010)。しかし、これらのツールのユーザビリティに関して、十分な検討がなされているとはいえない状況にある。

より本質的な困難さとして、「ブランドフリー問題」が挙げられる。Andersonは、その著書の最後で、ACT-R というブランド名を用いず、ACT-Rの研究コミュニティを超えたオーディエンスに対して、心の普遍的な説明を提供する必要性を説いた (Anderson, 2007)。この問題は、著者自身も、常々感じている。著者は、ACT-Rを強力で興味深いツールと考えているが、ACT-Rを褒め称えるために研究しているわけではない。しかし、ACT-Rを利用した研究を他者に伝える際に、ブランド名を出さざるを得ない。そして、この巨大なアーキテクチャをどう切り取って、普遍的な説明を構築すればよいのかと思悩むことになる。ただし、この問題はACT-Rに限らず、認知科学におけるモデルや理論の多くが保持している問題と考えられる。ブランドフリーな説明の確立には、認知の統合理論をさらに統合するメタな理論が必要になるのかもしれない。

謝辞

本稿に記された研究は、名古屋大学、北陸先端科学技術大学院大学、ペンシルバニア州立大学の多くの研究者との共同によって実施された。また、本稿の執筆にあたり、橋本敬氏より有益なコメントを受けた。ここに記して感謝申し上げます。

参考文献

Anderson, J. R. (2007). *How can the human mind occur in the physical universe?* New York: Oxford University Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, 111, 1036–1060.

Damasio, A. (1994). *Descartes' error: Emotion, rationality and the human brain*. New York: Putnam.

Dancy, C. L., Ritter, F., & Berry, K. (in press.). Towards adding a physiological substrate to ACT-R. In

Proceedings of BRIMS2012.

Gao, J., & Lee, J. D. (2006). Extending the decision field theory to model operators' reliance on automation in supervisory control situations. *IEEE Transactions on Systems Man and Cybernetics*, 36 (5), 943–959.

Konno, T., Morita, J., & Hashimoto, T. (2012). Symbol communication systems integrate implicit information in coordination tasks. In *Advances in Cognitive Neurodynamics (III)*. Springer.

Maehigashi, A., Miwa, K., Terai, H., Kojima, A., & Morita, J. (2011). Selection strategy of effort control: allocation of function to manual operator or automation system. In *Proceedings of 33rd annual conference of the cognitive science society* (pp. 1977-1982).

Morita, J., Konno, T., & Hashimoto, T. (submitted). The Role of imitation in generating a shared Communication System.

Morita, J., Miwa, K., Maehigashi, A., Terai, H., Kojima, K., & Ritter, F. (2011). Modeling decision making on the use of automation. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1971–1976).

Newell, A. (1994). *Unified theories of cognition*. Harvard University Press.

Newell, A., & Simon, H. A. (1976). Computer science as empirical inquiry: symbols and search. *Communications of the ACM*, 19(3), 113–126.

Reitter, D., & Lebiere, C. (2011a). How groups develop a specialized domain vocabulary: A cognitive multi-agent model. *Cognitive Systems Research*, 12, 175–185.

Reitter, D., & Lebiere, C. (2011b). Towards cognitive models of communication and group intelligence. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 734-739).

Ritter, F. (2009). *Two Cognitive Modeling Frontiers*. Information and Media Technologies.

Salvucci, D. (2006). Modeling driver behavior in a cognitive architecture. *Human Factors*, 48, 362-380.

Salvucci, D., & Taatgen, N. A. (2010). *The multitasking Mind*. New York: Oxford University Press.

Samsonovich, A. V. (2010). Toward a unified catalog of implemented cognitive architectures. In *Biologically Inspired Cognitive Architectures 2010* (pp. 195–244).

Stocco, A., Lebiere, C., & O'Reilly, R. (2010). The role of the basal ganglia–anterior prefrontal circuit as a biological instruction interpreter. In *Biologically Inspired Cognitive Architectures 2010* (pp. 153–162).

Tomasello, M. (1999). *The cultural origins of human cognition* (p. 256). Harvard University Press.

Zhao, C., Paik, J., Morgan, J. H., & Ritter, F. E. (2010). Validating a high level behavioral representation language (HERBAL): A docking study for ACT-R. In *Biologically Inspired Cognitive Architectures 2010* (pp. 181–188).

安西 祐一郎. (2011). *心と脳. 認知科学入門*. 岩波新書.