

2024.6 ブログ:「認知モデリングにおける記号処理アーキテクチャ」を読んで、の詳細
(→ <http://www.1968start.com/M/blog/index3.html#2406>)

「認知モデリングにおける記号処理アーキテクチャ」を読んで

中所武司

■このエッセイのきっかけ

下記の人工知能学会誌の解説論文について、第1次AIブームの1970年ころの私の修士論文「思考過程の数学的表現と模擬実験」の観点で読んでみた。

- ・「認知モデリングにおける記号処理アーキテクチャ」
人工知能, Vol. 39, No. 2, pp. 155-162 (March, 2024)

■内容の要約とコメント (→★)

(注) コメントの中で、以下の私の文献が引用されることがある。

- ・(注1) 修士論文「思考過程の数学的表現と模擬実験」(1971.3)
<https://www.1968start.com/M/bio/olduniv/shuuron.htm>
- ・(注2) 学会発表「思考過程のシミュレーション」、
電子通信学会オートマトン研究会資料、A70-76(1970.12)
<https://www.1968start.com/M/bio/olduniv/gakkai7012.html>

1. はじめに

- ・認知モデリングで利用されるコンピュータアーキテクチャには、「記号処理アーキテクチャ」と「ネットワークアーキテクチャ」があるが、後者が主流となった感がある。大規模言語モデル(LLM: Large Language Model)の出現により、伝統的記号処理AIは任務を終えた感がある。
- ・一方、認知モデリングは、人間の認知メカニズムや問題解決プロセスを研究する目的で、現在も、さまざまなタイプの記号処理アーキテクチャが用いられている。
本稿では、認知モデリングにおける記号処理アーキテクチャのもつ意味について述べる。

→★実用性ではLLMと比べられないが、私も興味を抱いている

「人間の認知メカニズムや問題解決プロセス」の探究には影響しないと思う。

2. 表象を巡る話題と認知アーキテクチャ

2.1 表象主義とアンチ表象主義

- ・認知モデリングで重要なのが、「**表象**」という概念である。
認知プロセスはこれらの**内部表象**に対する処理や計算として記述できる。
- ・古典的**表象主義**では、信念や知覚などの心的状態は、外界に対応する**内的表現**を含む。
これは、心的な絵やシンボルのようなもので、世界の物体や特性を反映する代替物である。
私達の心的プロセスは、これらの表象に依存して機能すると考えている。
これらの表象は、心的イメージ、シンボル、命題などさまざまな形をとる。
「心は、外的な対象、概念、または状態に対応する**内部表象**を構築し、操作する」と考える。

→★表象は、人間の思考モデル／思考過程の基本単位と考えてよいのかな？
表象に階層構造はあるのかな？

- ・この数十年、表象主義は、常にアンチ表象主義者の批判にさらされてきた。
アンチ表象主義においては、認知的活動は、本質的に、身体や環境が果たす役割が強調され、知的な行動は、表象の操作を必要とせずに達成される。
心の科学において、明示的な内部モデルは重要ではないと主張される。
意識の出現などの問題を、表象主義では適切に説明できないとしている。

→★内部モデルが必要ないということは、人間をブラックボックスとみるということ？

2.2 記号処理対ネットワーク処理

- ・初期の認知科学において、認知モデルといえば、基本的に記号処理モデルであった。
コンピュータ上で実行可能なプロダクションシステムやフレームシステムなどのアーキテクチャが開発され、それらの計算機資源を用いた研究に従事した。

→★1980年代の第2次AIブームのころの話かな。この時期は、知識表現が注目された。

- ・記号処理アーキテクチャは、「心は心的表象を処理・操作する」と考える表象主義を土台とし、その上での情報処理プロセスを、心が行う表象の処理のシミュレーションとみなした。

→★私の修論「思考過程の数学的表現と模擬実験」の模擬実験はシミュレーションのこと。
<参考：冒頭の（注1）>

- ・そのころ（1980年代半ば）に、ネットワークアーキテクチャが現れた。
このモデルでは、知能は、ネットワークのパターンの中に表現されていると捉え、
認知処理は、相互接続された要素間の動的相互作用の結果として遂行される。
このコネクショニズムは、表象主義に対して、いくつかの批判を展開した。
（表1に、両者の立場を、対比的に示す）

→★私が修論の研究をしていた1970年ころは、第1次AIブームの時で、パーセプトロンなど、ニューラルネットワークが注目されたが、パターン認識などに限界があり、下火になった。
それが1980年代半ばに再び脚光を浴びるようになった。

表1 表象主義・反表象主義の比較

項目	表象主義	反表象主義（コネクショニズム）
心的表象	モデル、ルール、記号などの内的心的表象の存在を強調する	明示的な内部表象の必要性に疑問を呈し、否定的である
処理	逐次的、直列的に処理される。 ここでの情報の操作とは、記号などの内的表象の操作を指す	情報は、並列的に処理される。 情報が複数のチャンネルや次元にわたって同時に処理される

アーキテクチャ	プロダクションシステムなどのルールベースアーキテクチャ, フレーム表現などの知識ベースアーキテクチャが用いられる	コネクショニストモデル, ニューラルネットワークなどの動的システムが代表的アーキテクチャである
AI との関係	エキスパートシステムなど, 古典的記号処理 AI と関連している	Big data を用いる Data centric approach やサブシンボリック AI と整合性が高い
身体性と環境	内的表象が重視され, 身体性や環境との相互作用も, 内部モデルをベースに記述される	身体や環境は, 認知主体の操作対象ではなく, 主体と一体化された認知活動の一部である
意図性	心的状態は, 外界や自分の思考の中にある対象に向けられ, 意図性 (intentionality) をもつ	意図性は, 主体と環境との身体化された相互作用から出現する。 身体感覚運動的相互作用は, 内的表象を必要とせず, 意味をもつ意図的な行動を生み出す
モジュール性	専門化された認知モジュールが内部表象を操作するという, 心のモジュール性を仮定	モジュール性に異議を唱え, 認知はより統合され, 複数のシステムに分散していると主張
構成性	表象は, 階層的, 構成的である。複雑な思考は, より単純な象徴的構成要素から構成される	文脈依存的で, 非記号的な処理を重視し, 構成的構造を仮定しない。
文脈や状況	文脈や状況は, 内部モデルが処理する対象である	文脈や状況は, 認知処理に直接統合され, 認知は文脈や状況に埋め込まれ, 両者は不可分である
学習	学習は, 内的な心的表象の獲得と精緻化の過程である	学習は, 外界との関わりを通して行われるネットワーク構造のチューニングである

→★表象主義と反表象主義 (コネクショニズム) に表1のような違いがあっても、いずれが優れているかという比較ではないと思う。表象主義は人間の認知機能を解明するサイエンスで、ニューラルネットワークは人間の脳の神経網に似せたネットワークで何ができるかというエンジニアリングなので、そもそも研究の目的が異なるのでは？

2.3 表象主義者の反撃

- コネクショニズムに反撃する哲学者は、人間の心は専門化された心的モジュールで構成され、それぞれのモジュールが特定の種類の情報を処理する役割を担っていると主張した。これらのモジュールは、心的表象を操作する。
- コネクショニズムに対する中心的攻撃は、「系統性の問題」である。系統性とは、認知システムが、構造化された行動を生み出す能力、例えば、概念や記号の新しい組合せをつくり出し、理解する能力のことである。言語や思考は、複雑な表現の意味を、その構成部分の意味から導き出す機能により成り立つ。

一方、知識を分節化しないコネクショニストモデルは、分散表現に依存しており、言語と思考における構成性を実現することが困難であると主張した。

→★人間の複雑な思考には、言語機能が不可欠と思うが、
両者は目的が異なるので、お互いに非難しあうことには違和感あり。

2.4 LLM の登場

・近年の大規模言語モデルの成功は、表象の实在について、根本的な疑問を投げかけている。これらのモデルの知能は、膨大な知識に基づくネットワークの中に織り込まれており、人間のような内部表現をもっていない可能性がある。その性能は、伝統的な表象に基づく認知というよりも、むしろパターン認識や、統計的関連付けの結果である。

→★大規模言語モデルが人間のような内部表現をもっていないのは、当たり前では？

・これらの AI は、人間に遜色ない複雑な会話を行い、心の研究者に大きな衝撃を与えた。ネットワーク上にある情報の組合せを超え、新たな価値を含むように見える回答もあった。これまで、創造性研究が対象とするプロデューサは、人間に限られてきた。が、ここに来て、そこに AI が加わるようになったのである。

→★「心の研究者に大きな衝撃を与えた」というのは、理解できない。
「新たな価値を含むように見える」としても、その判断は人間がしているので、LLMは、人間の道具の域を出ていないと思う。

・コネクショニズムに反撃する哲学者は、ネットワークアーキテクチャの弱点として、概念や記号の新しい組合せをつくり出し、理解する能力の欠如を指摘した。しかし、これらの LLM は、人間に遜色ないレベルで、その能力を有しているように見える。LLM は、コネクショニズムの弱点とされた系統性の課題を十分に克服しているように思える。

→★上記のコメントと同様で、この視点は理解できない。二者択一の話ではない。
人間の認知機能の仕組みの話と、特定のモデルやアーキテクチャの実用性の話は別物では？

3. 認知科学における記号処理アーキテクチャ

3.1 自覚的意識

・認知モデリングは、広い意味での心理学における研究アプローチの一つであり、特に概念や思考、問題解決を扱う認知心理学の目的は、人間の心の中で起こっていること、より具体的には心の中で情報がどのように処理されるのかを理解することである。

・AI などの人工物が「表象なし」で高度な知的機能を実現できるようになることと、心の科学における「表象」概念の重要さには、基本的に何の関係もない。私達人間は、現実には、表象を操作することによって「すごいこと」をしている。心の中で起こっていることを理解することは、表象の操作を理解することにほかならない。

→★上記のコメントと同様なので当然、同意。

前章までのような、両者を対立的に説明する必要はなかったのでは？

- ・私達は、分節化された内的経験として、自分が考えていることをモニタすることができるし、それを言語化することもできるので、明らかに表象を処理している自覚があるのだから、表象なしに、心で起こっていることを記述することは不適切である。

→★ここでの「自分が考えていることをモニタすることができる」という表現から、私の修論の最後で述べた以下の記述を思い出した：〈参考：冒頭の（注1）〉

<引用>

『思考は、人間の所有物でありながら、それは所有権なき所有であって、我々は、それを支配することはできない。これを何とか手中に収めたいという欲求にかられて、こうした研究に携わったのであるが、意に反し、振り返りにあった感が強い。それは、内観作業に頼りすぎたためかもしれない。』

しかし、行動主義心理学者のようにその客観性を重んじ、人間をブラックボックスと考え、その行動だけを研究対象とすることによって、ブラックボックスのメカニズムにこそ興味を持つ我々がえることのできる手がかりが不十分であるとすれば、やはり、内観も、一つの重要な手段としないわけにはいかない。

いずれにせよ、思考についての研究が、その思考によって行われるというジレンマから逃れることができないとすれば、我々はすでに第1歩からつまづいていることになる。研究主体が同時に研究対象であるという特異な分野の困難さをつくづくと感じた』

3.2 顕在処理と潜在処理

- ・認知心理学の領域では、認知処理を、顕在処理と潜在処理に分けて捉える。前者は、表象に基づく処理、後者は表象に基づかない処理に結び付いている。
- ・認知心理学は、人間の心の顕在処理に焦点を当ててきた。潜在処理は、顕在処理を支える副次的過程とされてきたが、近年、人間が行っている膨大な潜在処理の存在が明らかになるにつれて、潜在処理の重要性が強調されるようになった。極端な例では、潜在処理が主流で、顕在処理は、潜在処理が行った結果を、後付けで解釈しているだけであるという主張さえ、見られる。

→★私の修論で提案した思考過程のモデルは、これらの考えに似ている：

<参考：冒頭の（注2）>

潜在処理をニューラルネットワーク上の繰り返し処理で表現し、

一定の閾値を超えて出力された要素が言語化された顕在処理とみなした。これは、上記の「顕在処理は、潜在処理が行った結果を、後付けで解釈している」と似ている。

3.3 行動主義と認知主義

- ・行動主義心理学は、生体に対する刺激とその反応の関数関係を明らかにすることを目的とし、心をブラックボックスとみなすので、表象なし心理学であると言える。
- ・一方、認知心理学は、ブラックボックスの中で起こっていることを記述する言語構造の中に、表象という概念は織り込まれていると考えてよい。

→★私の修論（1971）の結びの引用（再掲）：

『行動主義心理学者のようにその客観性を重んじ、人間をブラックボックスと考え、その行動だけを研究対象とすることによって、ブラックボックスのメカニズムにこそ興味を持つ我々がえることのできる手がかりが不十分であるとすれば、やはり、内観も、一つの重要な手段としないわけにはいかない』

→★私の卒論（1969）では、以下の行動学の著書を参考にしている：

<参考文献> <https://www.1968start.com/M/bio/olduniv/soturon.htm>

15. D. O. Hebb : A textbook of psychology, W.B. Saunders company, 1958.
（白井ほか訳、行動学入門、紀伊国屋）

- ・認知心理学は、表象という概念を用いることで、人間の心の理解を、飛躍的に高めた。人間の認知の性質を利用した行動の誘導や、教育の分野における支援方法の開発、円滑なコミュニケーションをつくり出す仕掛け、認知的活動の支援や、インタフェースの開発に至るまで、認知心理学の成果が、さまざまな工学的、教育実践的応用に供されている。

4. 認知モデリングの教育実践

4.1 実験演習+認知モデリング

- ・著者らの認知科学の授業では、学生は、心理学の代表的な実験に従事し、心理学実験における知識、作法やスキルを、体験的に学習してゆく。そこで用いられるのが、記号処理アーキテクチャである。
- ・認知モデリングアプローチでは、理論に基づきモデルが構築され、シミュレーションによって、モデルの行動が確認される。実験的アプローチでは、実験の結果に基づいて、その背後にある理論の仮説が検証される。前者が、構成的（synthetic）、トップダウン的、演繹的（deductive）であるのに対して、後者は、分析的（analytic）、ボトムアップ的、帰納的（inductive）である。

→★私の修論（1971）は、ニューラルネットワークモデルを作成し、当時の大型コンピュータでシミュレーションしたので、前者に相当するかな。

4.2 DoCoPro アーキテクチャ

- ・記号処理アーキテクチャには、Formal logic, Schemabased, Network-based, Rule-based など、さまざまな知識表象を用いるタイプがある。著者らは、ルールベースの知識表現を用いるアーキテクチャを採用した。その理由は、ルールベースの知識表象の心理学的実在性が支持されていることや、

個々の知識に一定の独立性があること、知識のグレインサイズが直感に即していること、そこより個々のモデルの理解性が高いことなどである。

→★ルールベースは、第2次AIブームのときの主要な技術だった。

私も企業の研究所勤務時代にルールベースのエキスパートシステムを開発した経験がある。
(参考：学会発表)

- ・「マルチパラダイム型言語における対立概念の融合方式」、人工知能学会誌、4, 1, 77-87
(Jan. 1989) <https://www.1968start.com/M/p2/8901chuAIMulti>
- ・「エキスパートシステム構築ツール ES/X90 (1) :-開発思想」、情報処理学会第35回大会、3N-2, 1733-1734 (Sep. 1987) <https://www.1968start.com/M/p2/8709chuTaikai.pdf>

4.3 知識の外材化

- ・演習の参加者は、記号処理アーキテクチャベースの教育用プロダクションシステムでルールベースモデルの基礎を学び、マルチカラムの筆算の足し算の認知モデルを構築する。知識を書き下すことによって、自分が自動的に運用している知識に対して自覚的になる。

→★プロダクションシステムでは、知識をルールとして表現するので、わかりやすい。

5. 認知科学研究スキームの経験的学習

5.1 記憶の二重貯蔵モデル

- ・この演習では、認知科学の研究スキームを実践的に体験する例として、人間の記憶は、短期記憶と長期記憶という二つのサブシステムから構成されるという、記憶の二重貯蔵モデルを題材している。
- ・参加者は、時系列に提示される一連の単語を記憶し、その後、それらの単語を再生する。最初のほうに提示された項目と、最後のほうに提示された項目の再生率が、中盤に提示された項目の再生率に比して高くなるという現象が確認される。前者を初頭効果、後者を新近効果と呼ぶ。

5.2 実験とシミュレーション

- ・図3は、授業参加者による実験の結果と、認知モデルのシミュレーションの結果である。

5.3 Analogical comparison と Mental simulation

- ・参加者は、実験の結果とシミュレーションの結果を比較し、両者の共通性と差異性に注目し、その原因を探究する。これは、学習科学の領域での Analogical comparison に相当する。
- ・さらに、入力された単語列の短期記憶と長期記憶の相互作用がどのように行われるのかを、頭の中で追従する。これは、Mental simulation と呼ばれる認知活動である。

5.4 理論ベースの説明活動

- ・このような活動には、理論ベースの説明活動が促進されるという効果がある。実験結果の規則性やパターンについて、なぜそのようなパターンが現れるかが説明される。

→★A I の分野では、出力結果の妥当性を示すような説明が求められるが、
ニューラルネットワークでは、個々のニューロンと結合状態に戻づく説明は難しい。

6. まとめ

- 心理学における古くて新しい「表象」の話から始めて、認知モデリングにおける記号処理アーキテクチャ、ネットワークアーキテクチャの特性について述べた。近年の ChatGPT をはじめとする大規模言語モデルの出現においても、なお、人間の心を探求するための研究装置としての記号処理アーキテクチャの意義が顕在である。
- 認知心理学の授業実践では、心の中にある表象の操作という観点で、人間の心のはたらきを理解するために、記号処理アーキテクチャによる認知モデリングが重要な意味をもつ。
- 今後は、顕在意識と潜在意識、表象とパターン、記号処理とネットワーク処理が、相互に補完しながら、両者の相互作用を扱うアーキテクチャが主流になると思われるが、それぞれでしか明らかにすることができない心の機能があるという確信をもっている。

→★A I 応用の分野での実用性は、人間の認知機能の仕組みの話とは無関係と思う。
人間の認知機能の仕組みは、まだまだ未知の世界で興味深い。
単純な記憶でさえ、その保存やその想起の仕組みは解明されていないのでは？

以上