人工知能のこれまで (1)——記号的 AI

久木田水生

名古屋大学

名古屋経済大学「科学と人間社会 III」

そもそも人工知能って？

- マッカーシーによれば「人間がそれをやれば「知能を持っている」と思われるような振る舞いをする人工的なシステム」。
- 初期の人工知能は特に論理的な推論や言語を使うことに焦点を当てていた。

ジョン・マッカーシー
「人工知能」という言葉を作った。
By "null0" (https://www.flickr.com/photos/null0/272015955/)
[CC BY-SA 2.0 (https://creativecommons.org/licenses/by-sa/2.0)]

via Wikimedia Commons

この定義の困難

- 何が「知的」と見なされるかがかなり曖昧。
- 人間がやっても大して知的と思われなくても、人工知能の歴史において非常に大きな達成と見なされることがある。
- 例：猫の写っている画像とそうでない画像を見分ける。

人工知能の発展史

- 19 世紀末-1930 年代：記号論理学、計算理論の誕生
- 20 世紀半ば：コンピュータの誕生
- 1956 年：「人工知能」に関するダートマス会議
- 1950 年代：第一次人工知能ブーム。「論理的 AI」の時代。
- 1980 年代：第二次人工知能ブーム。「知識」、「身体化」、「コネクショニズム」などパラダイムの多様化。日本にもブームが波及。
- 2010 年代：第三次人工知能ブーム。「ビッグデータ」と「機械学習」の時代。インターネット上のビジネスに応用され大きな利益を生む。画像・映像処理、言語処理、ロボットなどと組み合わされることで応用の可能性が広がる。それと同時に社会への影響が懸念されるようになる。

論理学

- 厳密に正しい推論についての研究。
- 推論とはいくつかの前提（根拠）から結論を導き出す思考のこと。
- 正しい推論：前提がすべて正しい時には結論も正しい。
- 例：三段論法

  前提 *1*　すべての人間は死ぬ。
  前提 *2*　ソクラテスは人間である。

  結論　　ソクラテスは死ぬ。

  前提 *1*　すべての日本人は生魚を食べる。
  前提 *2*　日本人のノーベル平和賞受賞者がいる。

  結論　　生魚を食べるノーベル平和賞受賞者がいる。

- 三段論法は古代ギリシャのアリストテレスによって研究され、中世まで論理学の中心的なテーマだった。
- より一般的な推論の原理が探求されるのは 19 世紀になってから。

論理的推論の利点

- 論理的な推論を使えば、正しい前提からは必ず正しい結論が得られる。
- つまり確実に正しい前提だけから論理的推論を繰り返して得られた結論は正しいことが保証される。これが**証明する**ということ。
- 確実に正しいことが分かっていることだけから論理的推論だけを使えば、確実に正しい知識、すなわち真理が得られる。
- これが古代ギリシャにユークリッド幾何学において確立された数学の手法。
- 近代になりデカルトが哲学の方法論として重要視した。
- このように知識獲得の方法として論理的推論を重視する立場を**合理主義**という。

記号論理学

- 論理的に正しい推論一般を厳密に特徴づけることを目的に、19 世紀末から発展し、大きな成果を上げる。
- 推論の構造を明示化するために抽象的な記号によって推論を表現する。
- 限定的な範囲で、少数の原理からすべての論理的推論を導出することに成功する。
- 数学の一部分が論理学に還元されることを示す。→形式的数学、数学基礎論。
- それと同時に、数学には記号論理学の手法で厳密に特徴づけられない部分があることも明らかにした。

記号論理学における証明の例

公理 1 : $X \to (Y \to X)$
公理 2 : $(X \to (Y \to Z)) \to ((X \to Y) \to (X \to Z))$
公理 3 : $(\neg X \to Y) \to ((\neg X \to \neg Y) \to X)$
推論規則 : X と $X \to Y$ から Y を推論して良い.

$A \to A$ の証明

1. $(A \to ((A \to A) \to A))$（公理 1）
2. $(A \to ((A \to A) \to A)) \to ((A \to (A \to A)) \to (A \to A))$（公理 2）
3. $(A \to (A \to A)) \to (A \to A)$（1，2 と推論規則）
4. $A \to (A \to A)$（公理 1）
5. $A \to A$（3，4 と推論規則）

足し算の公理

1. 0 は自然数.
2. x が自然数ならば sx も自然数.
3. 任意の自然数 x, y について $(sx = sy \to x = y)$
4. 任意の自然数 x について $sx \neq 0$
5. 任意の自然数 x について $x + 0 = x$
6. 任意の自然数 x, y について $x + sy = s(x + y)$
7. 0 が性質 P を持ち，任意の自然数 x が P を持つことから sx もまた P を持つことが示されるならば，すべての自然数が P を持つ.

$x + y = y + x$ の証明

1. $0 + 0 = 0$.（公理 5）
2. $0 + x = x$ と仮定する.
3. $0 + sx = s(0 + x) = sx$.（公理 6 と 2）
4. 任意の自然数 x について $(0 + x = x)$ とすると $(0 + sx = sx)$ が言える.（2, 3 より）
5. すべての自然数 x について $0 + x = x$.（1，4 と公理 7 より）
6. すべての自然数 x について $0 + x = x + 0$.（5 と公理 5 より）
7. $sy + 0 = sy = s(y + 0)$.（公理 5）
8. $sy + x = s(y + x)$ と仮定する.
9. $sy + sx = s(sy + x) = ss(y + x) = s(y + sx)$.（公理 6 と 8）
10. 任意の自然数 x について $sy + x = s(y + x)$ とすると $sy + sx = s(y + sx)$ が言える.（8，9 より）
11. すべての自然数 x について $sy + x = s(y + x)$.（7，10 と公理 7 より）
12. $y + x = x + y$ と仮定する.
13. $sy + x = s(y + x) = s(x + y) = x + sy$.（12 と 11 より）
14. 任意の自然数 y について $y + x = x + y$ とすると $sy + x = x + sy$ が言える.（12，13 より）
15. 任意の自然数 y について $x + y = y + x$（6，14 と公理 7 より）

Intelligence とは？

- 推論や計算する
- 問題、パズルを解く
- 言葉を使う
- 計画を立てる。
- チェスなどのゲームをする
- 発見をする
- 創造をする
- などなど

ダートマス会議

- 1956 年、ダートマス大学にて開催。
- ジョン・マッカーシー、マービン・ミンスキー、ネイサン・ロチェスター、クロード・シャノンなどが中心となる。
- 初めて「人工知能（Artificial Intelligence）」という用語が大々的に使われた。

ダートマス会議での提案

本研究は、学習やその他のあらゆる知能の側面が原理的には、機械がそれをシミュレートできるくらい正確に記述できるという推測に基づいて進められる。どのようにすれば機械が言語を扱えるか、抽象を行い概念を形成できるか、現在は人間にしか扱えないような種類の問題を解けるか、そして自分自身を改良できるか。これらを発見するための試みが行われるだろう。ひと夏の間、慎重に選ばれた科学者のグループがともに取り組めば、これらの問題のいくつかにおいて重要な進展が成し遂げられると私たちは考えている。

J. McCarthy, M. L. Minsky, N. Rochester, C.E. Shannon, "A PROPOSAL FOR THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE", August 31, 1955.
`http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html`

Logic Theorist

- ダートマス会議で発表された、アレン・ニューウェル、ハーバート・サイモン、J・C・ショーが開発したコンピュータプログラム。
- 人間の論理的推論をシミュレートすることを意図して作られ、アルフレッド・ノース・ホワイトヘッドとバートランド・ラッセルの *Principia Mathematica*（『数学の原理』）の冒頭の 52 の定理のうち 38 を証明することができた。

記号的 AI、論理的 AI

- アレン・ニューウェルとハーバート・サイモンは「例えばコンピューターのような、物理的記号システムは知的な行動のための必要にして十分な手段を持っている」と主張した。
- A. Newell and H. Simon. "Computer science as empirical inquiry: Symbols and search." *Communications of the ACM*, 19(3):113–126, 1976.
- この考え方は「物理的記号システム仮説」と呼ばれる。
- 初期の人工知能においてはこのようなパラダイムが支配的であった。
- このようなパラダイムを**記号的 AI、論理的 AI** と呼ぶ。

シミュレーションと本物の思考

- コンピューターの理論的基礎を築いたアラン・チューリングは、振る舞い上人間と見分けがつかないほど流暢に人間とコミュニケーションができる機械は思考しているとみなすという考えを提唱した。
- これに対して哲学者のジョン・サールは、コンピューターは無意味な記号を操作しているだけで、その意味を理解してないから、本当に思考しているとは言えない、と反論した。
- 1990 年代以降、どのようにすれば記号の意味を理解できるかと言う問題は**記号接地問題**と呼ばれ、人工知能、ロボット工学において重要なテーマの一つになっている。

フレーム問題

- コンピュータは数学的な計算、論理的な推論、チェスなどのように、問題と答え（目標）、使える資源、考慮するべき要素が明確に定まっている課題を遂行するのはそれなりにうまくできる。
- また非常に単純化された現実のシミュレーションにおける課題（トイ・プロブレム）を遂行することもしばしば得意である。
- しかし現実の問題は必ずしもそうではない。

フレーム問題

- 現実の問題を解決しようとするとき、私たちは自分の行動がどのような副次的な帰結を持つかを考えなければならない。
- しかし現実世界においてある行為は極めて多くの帰結を持ちうる。
- 教卓を動かす→床が傷つく、教壇から机が落ちる、机が壊れる、黒板との距離が変わる、机の上のチョークが落ちる、パソコンが落ちる、ペンが落ちる、大きな音がする、教室の外の鳩が驚いて飛び立つ、木の枝が揺れる、などなど。
- 論理的には、一つの前提は無限の多くの帰結を持ちうる。
- 人間はそういった帰結のうちのいくつかを無意識のうちに選んで考察している。
- しかしコンピューターにそのような選択をさせることは極めて難しい。

論理的AIの限界

- 記号論理学をベースとした推論に基づいた人工知能では現実の問題に対処することは難しい、ということが認識されるようになる。
- 例えばコンピューターには状況や文脈を理解することや、人間にとって当たり前の常識を前提して思考することが致命的に苦手。
- テリー・ウィノグラードの例：「市評議会はデモ隊にデモの許可を出さなかった。彼らは暴動を【恐れていた／辞さない姿勢だった】からである」
- ここでの「彼ら」の指す対象は【】内の語句に応じて変化する。
- ある程度の常識のある人間ならば「彼ら」が何を指すかはすぐに理解できるが、コンピューターにはこれが極めて難しい。

論理的AIの限界

- コンピューターは、人間の知能をシミュレートして、人間の代わりに課題を遂行するのではなく、あくまでも人間の知能をサポートするものとして使うことが提案される。
- Cf. テリー・ウィノグラード、フェルナンド・フローレス、『コンピュータと認知を理解する――人工知能の限界と新しい設計理念』、平賀譲訳、産業図書、1987。

ドレイファスによる批判

- 哲学者のヒューバート・ドレイファスは、1964 年にランド研究所に提出した報告書「錬金術と人工知能」において、人工知能研究者が実現できないことを吹聴して研究資金を詐取している、と強く非難。
- 人間の知能においては経験によって培われる勘、体に染みついた感覚、暗黙の知識などが重要な役割を果たす。
- こういったものは論理的な規則のようなもので表現することは不可能であり、したがって機械は真の意味で知能を実現することはできない、と主張した。