

情報の「理解」の定式化に向けて

Toward mathematical formulation of what understanding is

日高 昇平[†]

Shohei Hidaka

[†]北陸先端科学技術大学院大学

Japan Advanced Institute of Science and Technology

shhidaka@jaist.ac.jp

Abstract

A long-standing dream in research on artificial intelligence (AI) is to build a *strong* AI termed by Searle (1980), which understands and processes the input, unlike a *weak* AI which just processes it as programmed. The key question here is what understanding is. In this study, I propose a direction toward a mathematical formulation of understanding. Revisiting Searle's Chinese room argument, we find the importance of information transmission, and I propose an extension of Shannon's information theory by relaxing its key assumption on the common codebook. Further, I point out the essential difference between learning and understanding. From this series of arguments, I summarize it by showing research steps toward a formulation of understanding.

Keywords — Understanding, strong artificial intelligence, information theory, and learning.

1. 理解と強いAI

人工知能(以下 AI)の研究における哲学は、大きく二つに分けられる。一つの流儀は、ある特定の機構をもった機械あるいは計算プログラムを作り、人あるいはその他動物の行動や心の働きとその機械を比較し、人の心の働きと同種の機構を実現する一例として機械をみなす方法論である。もう一つは、人を含め「知能」をもつものは、特定の機能の実現によって定義できるとし、逆にそうした機能実現こそが「知能」の本質であると考えた方法論である。前者は特定の認知機能を実現するモデルとしての AI、後者は機能が実現できる限り人と機械を区別せず、高い機能をもった AI により人と同等の「知能」を人工的に実現できると考える。Searle(1980)は前者を「弱い AI」、後者を「強い AI」と呼び、強い AI の思想に懐疑的な批判を行った。

Searle の強い AI に対する批判は、「中国語の部屋」と呼ばれる思考実験に集約される。この思考実験では、中国語を知らない Searle 自身が、小窓からスクリプトの出し入れが可能である部屋に入り、彼にとっては理解できない中国語の問いを受け取り、Searle の母国語

である英語で書かれた回答手順書にしたがって中国語の答えを返す(図 1(A))。もし英語で書かれた手順書が十分に洗練されていれば、Searle 自身が中国語をまったく理解できないとしても、部屋から出てくる中国語の答えは、外の観察者からは中国語話者と違いが判別できないであろう。したがって、この中国語の部屋における、英語の手順書(プログラム)、そしてそれを実現する Searle 自身(機械・AI)は中国語を理解できずに、言語的な応答能力によって知能を定義する Turing test に合格するはずである。

以上の中国語の部屋論法により、Searle は強い AI と弱い AI をわける鍵が「理解」とし、単に特定の機能あるいは入出力の対応関係の実現だけにとどまらない「理解」なるものが知能の本質であると批判した。しかし、Searle 自身は「理解」が何であるか明確な定義を与えておらず、直感的な事例の列挙による説明にとどめている。

2. マネ碁と情報の伝達

本稿では、「理解」とは何かという問いに対し、直観や共感に訴ることなく定式化を与えたい。まず直観に訴えるために作られた中国語の部屋論法を、その本質を失わずより簡潔にした思考実験を考える。

ごく最近、チェスや将棋より複雑な探索空間をもつと考えられてきた囲碁でも、ついに DeepMind の開発した Alpha Go が人間のトッププレイヤーを破ることとなった(Silver et al., 2016)。この囲碁を例にとり、名人と同等以上の強さをもつ AI プログラムが作れたなら囲碁版の Turing test に合格と認めるとしよう。それには、AlphaGo の開発を待たずとも以下のようなごく簡単なプログラム(マネ碁 AI)を作ればよい(図 1(B))。

1. 対局 1 では名人と AI プレイヤーが、対局 2 では名人と肩を並べる挑戦者と AI プレイヤーが対局する。

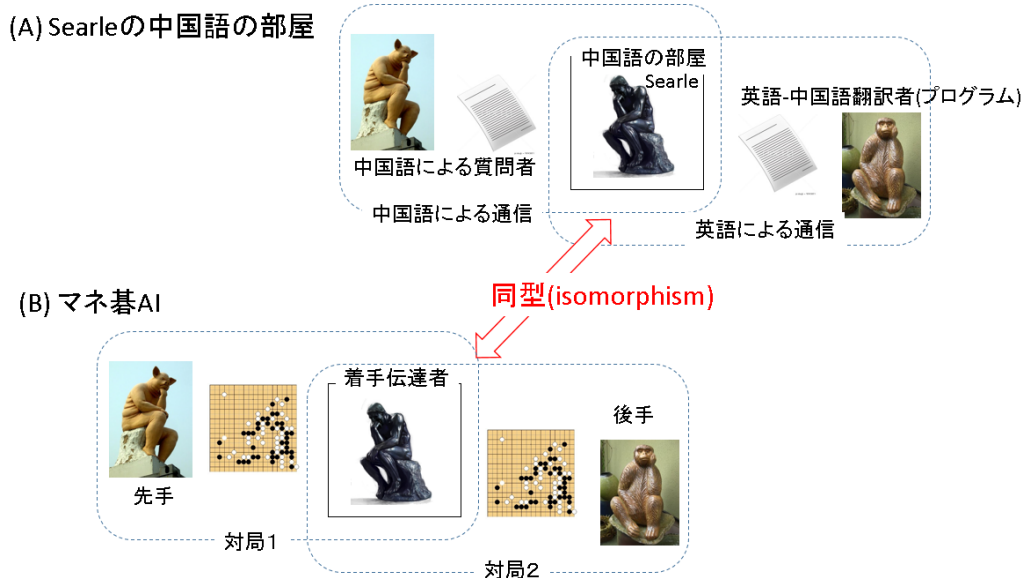


図 1 : (A)Searle の中国語の部屋と(B)マネ碁 AI(着手伝達エージェント)の同型性

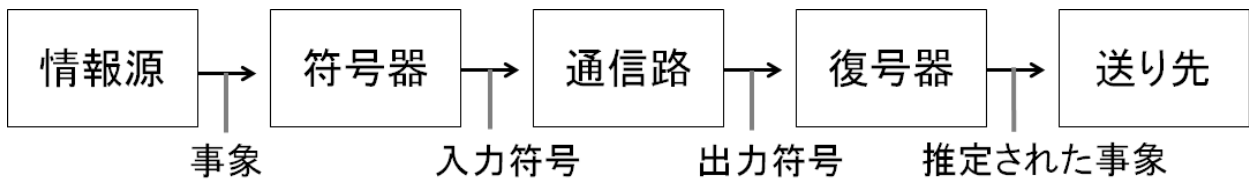


図 2: Shannon の情報伝達モデル。送り手(情報源)から送り先へ、通信路を通じて符号を送ることによって信号を伝える。一般に通信路は確率的誤りを含むが、信号を符号へと変換する符号器、符号を信号へと変換する復号器を適切に設計することで、誤り確率の任意に小さい通信が可能である。この Shannon モデルでは、復号器の設計には符号器の知識を必要とする。

2. 対局 1 では AI プレイヤーが後手番となり、対局 2 では AI プレイヤーが先手番とする。
3. 対局 1 で名人の第一手をまち、それと同じ手を AI プレイヤーは対局 2 で打つ。
4. 対局 2 での挑戦者の第二手をまち、それと同じ手を AI プレイヤーは対局 1 で打つ。

以下これを繰り返すと、AI プレイヤーは二局とも引き分けあるいは、一勝一敗の結果を得る。従って、この AI プレイヤーは名人と同格であり、囲碁版の Turing test (Turing, 1950) を合格する。

このマネ碁の思考実験は、Searle 論法の中国語話者を名人、また Searle に英語の手順書を与える翻訳者を挑戦者とする、本質的に中国語の部屋論法と同型である(図 1)。一方、Searle の「理解」の直観的説明なしに、このマネ碁 AI が囲碁の理解を必要としないことが

明白である。つまり、このマネ碁 AI は、対局 1 と 2 の間の通信あるいは情報の伝達しているに過ぎない。これは中国語の部屋論法において、Searle が中国語話者と翻訳者の問答を媒介するのと同型である。翻って、中国語の部屋論法は、本質的に情報の伝達と理解の違いを指摘しようとしているのである。

3. 情報の伝達を超えて

以上の思考実験から、「理解」と情報伝達の類似性および、本質的な違いが示された。本論では、情報伝達の数理を再考し、情報「理解」の定式化に向けた提言を行いたい。

Shannon (1948)によって定式化された情報理論は、伝達可能性を与える通信容量定理により、情報

の概念に明確な定義を与える。この情報理論の定式化において、Shannon は情報の「理解」という問題の定式化を巧妙な前提を置くことにより避けて、情報の「伝達」のみに焦点絞って数学的に表現している。逆に言えば、情報理論の前提の再考および緩和により、「理解」の定式化できる。実際、Shannon と Weaver (1948) は、明確に情報の伝達・理解・因果性の3つを区別し、情報理論によって伝達の問題だけを解くことを宣言している。

Shannon の情報伝達モデル(図 2)の再考から得られる洞察は、伝達される符号列と元の信号との対応表である符号器が、情報の送り手と受け手の間で**事前に共有されているという仮定(共有符号器の仮定)**が、この定式化において重要な役割を果たしているという点である。この仮定により、最も重要な伝えられた符号列の「解釈」あるいは「理解」の問題に悩まされることなく符号列が誤りなく伝達可能な条件にのみ焦点を絞って数学的な議論を展開できる(詳細については、日高(2016; 査読中)および Hidaka & Kashyap (2014)を参照)。

4. 情報・学習・理解

一般に、学習と理解は異なると考えられる(たとえば、佐伯(2007))。たとえば、算数教育において、特定の手続きで計算ができるようになるには、繰り返し計算問題を解くことで成績が上がる(学習)。しかし、これは、応用的な問いに対する答えから察せられる本質的な数学的な理解とは必ずしも同じではない(銀林, 2007)。この意味で、情報伝達と学習、さらに学習と理解と違いは考察に値する。

情報の伝達では、静的な復号器を想定していたが、データから復号器を近似的に推定する方法のひとつとして学習について考えよう。情報理論の枠組み(図 2)において、「学習」とは、近似的に復号器を再構成することである。一例として、教師あり学習による分類問題を解くことを考えよう。この場合、ある特徴空間上に表現された事例(データ/符号)の少なくとも一部に対して同一性を明示する教師信号が与えられる。教師信号は事例のクラスを明示し、同じクラスに属する事例を同じクラスに分類し、異なるクラスに属する事例を異なるクラスに分類する分類器は、符号(事例の特徴)から信号(クラス)を与える写像(復号器)の近似とみなせる。この問題において、特定の学習則により、近似的な復号器を構成可能であれば、

一見、この学習により前述のメタ復号器の問題を解けるように見える。しかし、こうした教師あり分類学習は、符号、信号の両方が与えられており、メタ復号器の問題の一部にトップダウンに正解を与えている。従って、符号・信号の両方が与えられた条件下での復号器の再構成は、それだけではメタ復号器の問題の十分な解とはいえない。

これは他の種類の学習においても同様である。次の例として、教師なし学習のひとつとしてクラスタリングを考えてみよう。一般に、クラスタリングでは、事例の特徴と、その特徴空間上の類似性(距離関数)が与えられ、クラスタ間と比較してクラスタ内の事例の類似度が高くなるように事例のクラス分けを行う。この種の問題では、信号(クラス教師信号)は陽に与えられていない。しかし、特徴空間の距離関数が与えられているため、暗に復号器に相当する類似性が与えられている。この意味で、やはり教師なし学習もメタ復号器の問題の解にはなりえない。

ここで述べたように、特徴の解釈たる類似性が与えられれば分類が可能であり(教師なし学習)、あるいは、クラス信号が与えられれば、そのクラスを与えるのに適した類似度の関数を構成できる(教師あり学習)。しかし、その両方を同時に解決することは難しい。こうしたジレンマは、概念獲得の根本的な問題であると考えられている(Murphy & Medin, 1985)。

近年の機械学習分野の発展により、特徴とクラスの繰り返し相互に推定する方法(e.g., Expectation-Maximization(EM)法(Dempster et al., 1977)など)など、こうした特徴とクラスの相互参照性を部分的に解決する方法論は存在する。しかし、こうした解決法は、復号器の推定と信号の推定の両面で使える目的関数(上記 EM 法の場合、尤度関数)を与えており、本質的には与えられた枠組みの中で部分的な変数を統計的に推定する構造となっている。従って、本質的に、「学習」は与えられた条件下で特定の写像(復号器、あるいは信号)を構成する計算とみなせる。

すでに述べたとおり、学習と理解が異なるとすれば、こうした学習において本質的にみられる、符号・復号器・信号の相互依存的な関係を解消する性質を「理解」が持つと考えられる。この学習と理解の違いは、それぞれ「(部分的な)記述」と「説明」ではないだろうか。これらの本質的な違いは、「記述」が断

片的な情報の集合であるのに対し、「説明」がその**局所的記述と全体の整合性を有する情報の集合**である点にある。「学習」が、与えられた他の部分を補うように、未知の部分を推定する(記述する)のに対し、「理解」は、完結した「説明」を構成するのに対する述語と考えられるのではないだろうか。

5. 理解の定式化に向けて

以上をまとめて本論では、この共有符号器の仮定を緩和することで、情報の理解の数学的な定式化を模索することを提案する。

結論として、本研究では以下の理論的な予想・研究プログラムを提案する。

(1) 情報の理解とは、信号から符号への写像である符号器を、符号列だけから推定する復号器を、推定することである。陽には与えられない復号器を、符号列のみから推定するこの問題を**メタ復号器の問題**と呼ぶ。

(2) 情報の伝達と理解に対応する「記述」と「説明」の本質的な違いを捉え、全体的な整合性を有する情報の単位を同定する問題を**全体性単位の問題**と呼ぶ。

個々の記述から全体性へと対応付ける(～説明する)ことが(1)メタ復号化であり、またメタ復号化に過不足のない情報の単位が、(2)全体性の単位であると言える予想される。従って、この2つの相補的な問題を追究する研究プログラムの推進により、情報理解の数学的な定式化を提案する。(1)および

(2)は、それぞれ Shannon の情報理論の、共有符号器のない状況への拡張、および多数の送り手と受け手の全体的な整合性を制約としておいた通信への拡張、に相当する。大会発表では、著者のこれまでの研究(日高, 2016; 査読中; Hidaka & Kashyap, 2014)および、最新の研究(Hidaka, in preparation)を具体的な基礎として、研究の方向を論じたい。

謝辞

本研究は平成28年度科学研究費補助金若手研究A 16H05860 の助成を受けて行われた。

参考文献

[1] Dempster, A. P., Laird, N. M., & Rubin, D. B.

(1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, 1-38.

[2] Hidaka, S. and Kashyap, N. (2014). The Generalist Approach to Frame Problems, in Proceedings of The Third Asian Conference on Information Systems, pp. 318–325.

[3] Hidaka, S. (in preparation). Unbiased and asymptotically optimal estimator for partition with the minimal information loss.

[4] Murphy, G. L., & Medin, D. L. (1985). The role of theories in conceptual coherence. *Psychological review*, 92(3), 289-316.

[5] Searle, J. R. (1980). Minds, brains, and programs, Behavioral and brain sciences, Vol. 3, No. 03, pp. 417–424.

[6] Shannon, C. E. (1948). A Mathematical Theory of Communication, Bell System Technical Journal, Vol. 27, pp. 379–423.

[7] Shannon, C. E. and Weaver, W. (1949). The mathematical theory of communication, University of Illinois press.

[8] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Driessche, van den G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search, Nature, Vol. 529, No. 7587, pp. 484-489 (2016)

[9] Turing, A. M. (1950). Computing machinery and intelligence, Mind, pp. 433–460.

[10] 佐伯 胖 (Ed.). (2007). *理解とは何か* (Vol. 2). 東京大学出版会.

[11] 銀林浩. (2007). 算数・数学における理解 佐伯胖 (編) コレクション認知科学 2 理解とは何か 新装版 東京大学出版会.

[11] 日高 昇平 (2016). 情報理論から意味理解の理論へ., 第 30 回人工知能学会全国大会.

[12] 日高 昇平 (査読中). 情報の伝達から理解へ.