

「人間知能」のモデル化におけるAI(人工知能) 特に言語モデルGPT-3の位置付け

水野 義之

京都女子大学

mizuno@kyoto-wu.ac.jp

Modeling of human intelligence and its relation to an NLP model of GPT-3

Yoshiyuki Mizuno

Kyoto Women's University

概要

大学における一般情報教育の中で、人工知能 (AI) を適切に扱うためには、そもそもの「情報」の再定義に加えて、そもそも「人間の知能」とはどういうものかに関する再定義が必要であると考えられる。なぜなら人間は、自分の知能を理解しているとは考え難いからである。著者はこの視点から近年、「情報」の再定義（「情報の発展モデル」）を基礎として、「人間知能」のモデル化を提唱している（「情報と知能の発展モデル」）。そこでこの報告では、特に近年の自動翻訳の劇的な進歩など、2017年以降の発展著しい自然言語処理に着目し、その中でも特に、一見して人間の能力と区別が付かないかに見える言語モデルとして知られるGPT-3が、著者の提案する「情報と知能の発展モデル」の中に、どのように位置づけられるかについて議論する。

1 はじめに

この報告では、著者の提唱する「人間知能モデル」の中に、自然言語処理のAIで知られるGPT-3を位置づけることを試みる。これを行う理由は、GPT-3を人間が初めて見た時に受ける衝撃の大きさは無視できないと判断したためである。

大学における一般情報教育（例えば文献[1]の第11章「人工知能(AI)とデータ科学」）の中で、人工知能 (AI) を適切に扱うためには、「情報」の再定義に加えて「人間知能」に関する再定義が必要である。著者はこの論点から近年、「情報」の再定義の提唱である「情報の発展モデル」を基礎として、「人間知能」のモデル化を提案している。後者を「情報と知能の発展モデル」と呼ぶ。

しかし近年の自動翻訳の劇的な進歩など、2017年以降に大きく発展した自然言語処理、特にその中でも人間の発話と区別が付かないレベルの

処理ができる言語モデルとして知られるGPT-3 [2]の位置付けが問題となると考えられる。そこでこの論考では、このGPT-3が、著者の提案する「情報と知能の発展モデル」の中にどのように位置づけられるかについて議論する。

2 人間知能のモデル化 (1)

2.1 大学情報教育との関係：情報の発展モデル

大学における一般情報教育は、その出発点として、そもそも「情報」とは何かという問題を取り上げ、「情報」そのものについて深く考えさせることから始めることが多い[1]。この問題を考えることで初めて、データと情報の区別、情報と知識の区別、その先にある創発的な「知恵」、という4つの対象（データ、情報、知識、知恵）を位置づけることができると考えられるからである。

この理解の上で、その先には何があるかとい

う問題を考えよう。それは何かの総合（統合）であるに違いない。なぜならバラバラでも知恵は知恵であるからである。これで知識生産の流れは5段階となって、データ、情報、知識、知恵、統合、である。しかしこれで終わりではなく、その統合的な存在をよりよく理解するために「分析」が必要となると考える。その分析こそが「方法としての科学」である。この「方法としての科学」は、「データ」に始まる一連の流れとは逆向きであり、科学は再度、データを生み出す。

これに対してデータ、情報、知識、知恵、統合という5つの対象の「流れ」は「知識としての科学」、つまり帰納的な科学知識の生産の過程の流れである。従って知識生産の過程と、データ生成の過程は、最終的には螺旋的に発展するはずである。そういうモデルを著者は提案している。これを「情報の発展モデル」と呼ぶ（図1）[3]。

「情報の発展モデル」 (水野、2009)

出典：水野義之, 2009, 「情報社会における「情報」の発展モデル」
『日本社会情報学会第24回全国大会研究発表論文集』:184-187.

データ、情報、知識、知恵、芸術、科学

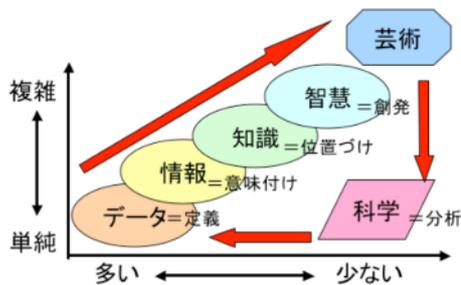


図1：初期の「情報の発展モデル」[3]

2.2 AI時代の謎：情報と知能の関係

これはこれで、良かった時期があった。しかし2016年の「AlphaGo事件」を経て、人工知能（AI）の時代になった。そして考えた。この図の中で、「知能」はどこにあるのか。

そこで考えたのが、知能の成分分解（図2）である。著者は「知能」には4成分あると提唱している。知能とは、知性、理性、感性、悟性、この4つであるとする。すなわちデータを情報にする段階で必要になる「知能」は「知性」である。

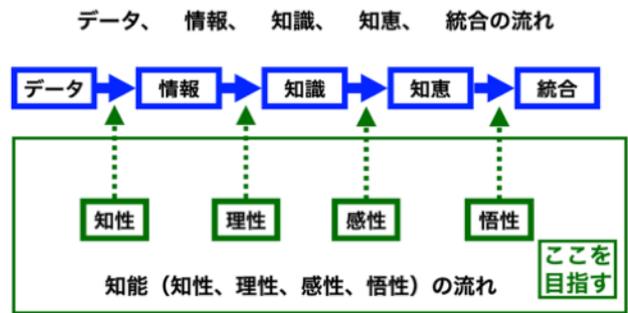


図2：「情報の発展モデル」（図1）に対応して、人間知能は、図中の4成分に分解できる。

情報を知識にする段階で必要になる「知能」は「理性」である。知識から知恵を創発する段階で使う「知能」は「感性」である。最後に種々の知恵を総合して1つの「統合」を創造する段階で必要になる能力を「悟性」と名付ける。

このようにして「情報の流れ」と「知能の成分」との対応関係（図2）を発見した[4-7]。

ここでさらに、これに双方向性を加える。なぜなら、知識生産の流れには、データから統合に向かう帰納の流れ（「知識としての科学」と、統合からデータに向かう演繹の流れ（「方法としての科学」）の両方があるからである。

このようにして得られたのが、上下の流れを一体化させた双方向の流れ（図3）である。この対応付けによって、「デカルトの理性」も「カントの認識論」も1つの図の中で、的確に対応することも示される[5-7]。

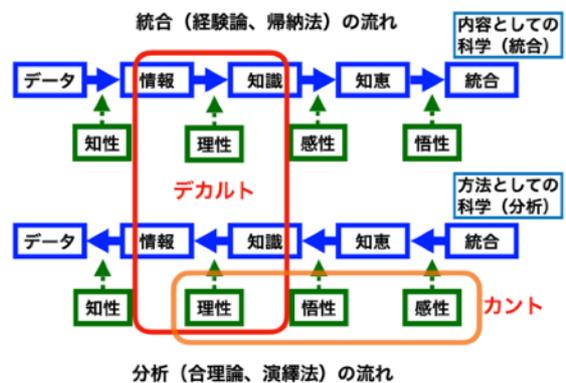


図3：「人間知能のモデル」と、歴史的に知られる2つの知能モデル（デカルトの理性論とカントの認識論）との比較。

この対応の発見は興味深いですが、実は何の証明にもならない。しかし、ここで提唱した統合的なモデル（図3）が、モデル自体の内部で整合的であることは容易に了解される。この事実は重要である。

2.3 情報と知能のモデルにおけるAIの発見

このような考察を経て、図1と図3を組み合わせた発展形を考えることができる。そして最終的に図4（下記）が得られた。これが著者が提唱している「情報と知能の発展モデル」である。

この図4の特徴として、図の左下の領域（「データ」が「情報」になる段階の領域の一部）において、2016年（AlphaGoの時代）までに我々が知っていたところの「AIの得意な領域」を、適切に位置づけることが出来ていると、著者は考える。

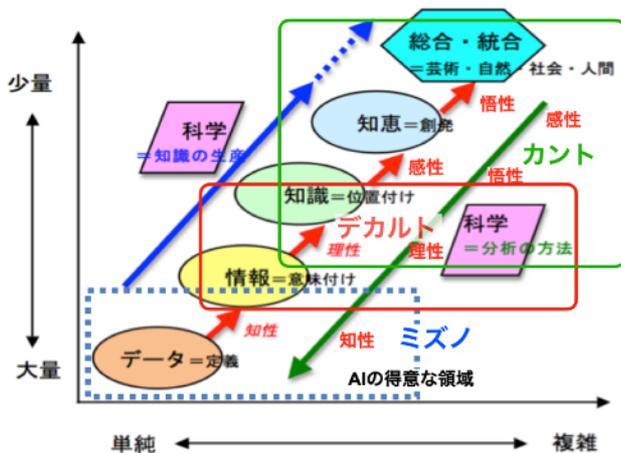


図4：「情報と知能の発展モデル」。これは「情報の発展モデル」（図1）の発展形であり、人間知能の展開モデル（図3）を、図1内部に組み込んだモデル。

すなわち、2016年3月のAlphaGoの登場で世界中が衝撃を受けたのは、いわゆるディープラーニング（深層学習、「教師あり機械学習」）と、強化学習、そしてモンテカルロ・ツリー探索という、3つの要素技術の統合としてのAlphaGoであり、その総合性能の高さによってであった。

しかしこれは実は、囲碁というゲームにおける「盤面評価」や「次の一手評価」など、全ては

画像認識技術と一般的探索技術の延長としての技術開発に過ぎなかった、とも言えるのである。もちろんAlphaGo事件が世界に与えた影響は巨大かつ甚大で、それ自体が深刻だった。例えば、この対戦で敗れた人間側の李世石氏は結局、囲碁棋士を引退してしまった。

しかし、そうこうするうちに2017年以降、人間社会はAI技術の次の発展に、再び衝撃を受けることになる。それが自然言語処理技術の発展であった。

2.4 自然言語処理AIが与えた衝撃

自然言語処理における3つの出来事に言及する。第1に、2018年暮からGoogle自動翻訳の性能が急に上がったが、背後にBERT（Bidirectional Encoder Representations from Transformers）[8]と呼ばれた技術的発展があったこと。

第2に、2019年にOpenAIが開発した自然言語処理 GPT-2 [9] の「物語作成」の能力が高く、ソフトウェアを非公開にするとの発表があったこと（公開のOpenAIが作成した技術にもかかわらず、フェイクニュースへの悪用が心配され、非公開になった。その事実が、社会に衝撃を与えた）。

そして第3にGPT-3の登場だ。世界中の英語版 Wikipedhiaの全文を記憶し、その膨大な例文空間をコード化した上で、適切な文を生成し、対話でも「人間」かと思うレベルで、自然言語を「話す」ことができた。

GPT-3 [2] の衝撃は、素人にも専門家にも同様に生じるようである。例えば2020年度の米国MITのFridman教授の授業[10]では、GPT-3によって哲学者が、理解とは何かとの問いを発するようになったと語られる。

3 人間知能のモデル化（2）

3.1 自然言語処理AIのGPT-3の技術

前節で言及した自然言語処理GPT-3で使われる技術は、BERTやGPT-2に使われたものと同じであ

る。GPT-3ではGPT-2の約100倍のデータ量を学習させており、精度はさらに劇的に向上した。これが単純にスケールした事実は開発者を驚かせた。それは、まだ伸びることを意味するからである。これらの技術にはAttentionとTransformerと呼ばれる仕組み [11] が共通している。

この原理は、比較的単純であることがポイントだ。ここでは単語の辞書対応を使ってベクトル化した上で、そのエンコード・デコードのプロセスの並列化と組合せ、繰り返しと多重化という比較的単純な線形変換の処理を工夫して単語予測モデルを構築し、そのパラメータの訓練用にはWikipedia上の多種多様な自然言語の文章（あるいはプログラミングのソースコード、数値・数式などを含む）を使い、その予測精度を上げるという意味の機械学習をさせている。

分かりやすくするため「擬人化」した言い方で言い換えれば、例題を大量に覚えて、それを出来るだけ上手に真似るという方法で学び、聞かれたことには、すでに学んだパターンに対応させて、対応する中で最適な答えを返す、というだけである。しかしこれはよく考えると、一般的な人間が、頭の中でやっていることと、ほとんど同じであると考えすることはできないだろうか。

3.2 自然言語処理GPT-3の技術の意味

次に、GPT-3のような高精度の自然言語処理が人間精神や人間知能において、持つ意味を考える。

人間は言語を使う時、あるいは言語を学ぶ時に、自分の頭の中で一体、何をやっているのか。自分自身では知る方法を持たない。しかし人間のセンサーとしての感覚やプロセッサとしての脳的能力は、有限の精度しか持たない。このため普通の人間は、その精度の範囲でおおよそ同じであれば、それを同一と見做しているのではないだろうか。

例えば、デジタル写真であろうがフィルム写真であろうが、ある程度の高精細になれば、そのピクセル一個一個は人間には認識できない。このため、ある距離で眺めて写真としての違いが分からなければ（つまり十分に似ていれば）、人間は両

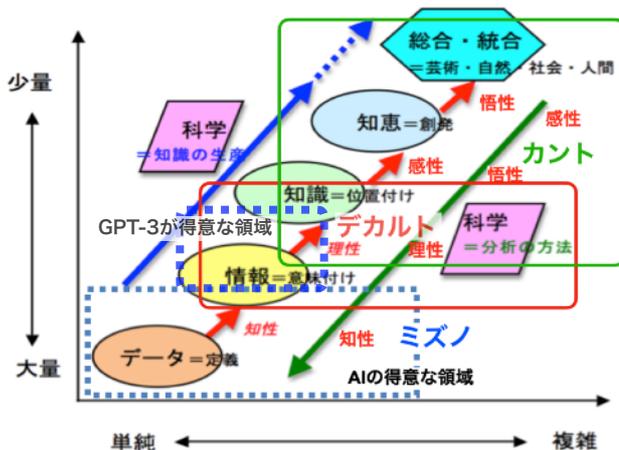


図5：「情報と知能の発展モデル2」。

図4の「情報と知能の発展モデル」の中に「GPT-3が得意な領域」と書き込んだ。

者を同じであると見做すであろう。これは自然である。

空を飛ぶという機能でも、人間の目的の枠内で飛行機と鳥とが、同じく飛ぶという機能を果たせれば、両者は全く同じではない（全然違う）としても、それは十分に人間の役に立つと、人間は考えるであろうと思われる。

ここで指摘している類似性の持つ同一性は、まさにアナログ（あるいはanalogous）の意味そのものである。

もう一つ例を挙げる。人間の遺伝現象はデジタル現象であることは、DNA二重螺旋がわずか4種類の塩基配列のデジタルパターンでできていることから自明だ。それにもかかわらず、その遺伝現象は人間の感覚では、似ていることと理解されていると思われる。親子が似ている時、ゲノム配列や特定の遺伝子の染色体での場所を思い出す人は余りいない。我々の何かの理解とは、類似性の枠内で起こる現象である。つまり遺伝現象でも、デジタル機構がもたらすアナログ現象を、我々は見ている。元々がデジタルデータであるかアナログデータであるかには関係しない。

何かの意味に意味があるのはなぜか、という問題を考える。我々が何かの言葉（言語的認識）や画像（視覚的認識）に意味を感じるのは、我々の脳内には、ある種のモデルがあつて、それとの

一致度が高いからであると言い換えることが出来る。我々人間は、自分の脳内モデルを外界に投影している。それによって世界をみて、それを解釈し、そこに意味を見出しているのである。

つまり我々から見て同じであれば、それは同じであると判断している事になる。そもそも、2つのものが同じ、とはどういうことかと考えると、それは厳密にはあり得ないことがわかる。なぜなら元々2つの別のことだからである。

機械翻訳も言語能力も同じである。我々が言葉で対話できれば、それで良いと考えられる。もちろん、おそらく機械（あるいはAI）自体は意味を理解していない。それはよく指摘される通りである。確かに機械自体は意味を理解していないだろう。しかしそれは人間が機械に、意味の意味を教えていないのだから、当然のことだ。

ここでの考察の結論は、自然言語処理GPT-3のような高精度の言語処理技術が出現した現在において、その機械（AI）の能力は、人間の能力の一部を置き換えることができると、判断できるということだ。

3.3 「人間知能」におけるGPT-3の位置付け

ここでいうGPT-3のような人間に相当するレベルの言語処理の能力は、著者が提唱している「情報と知能の発展モデル」（図4）に戻って考えると、一体、どこに相当するのであろうか。

それはおそらく、既存の画像認識AIのようなデータ→情報の間の能力、すなわち認知あるいは認識（著者の言葉では「知性」）の部分、明らかに超えているのではないだろうか。なぜなら知性は、データに情報を見出すために使う知能であり、バラバラの情報でも良かったからである。

2021年現在のGPT-3などAIが持つ言語処理の持つ知能は、おそらく、情報→知識の間を変換する知能であるところの「理性」の一部であると考えられる。なぜなら現今のAIは単語間の適切な繋がり判断や、人間が意味を感じられる文章を機械的に生成できるからである。そしてこのような情報の構造化こそ、著者が知識（理性の働き）と呼ぶものに他ならないからである。この意味で、こ

のGPT-3レベルの言語処理能力は、ここでいう知能の中の「理性」の獲得に対応させられる（AIが置き換えうる）と思われる。

言い換えれば、2017年のAttentionやTransformerの歴史的登場 [11] 以来、2021年現在に至る自然言語処理能力の劇的発展によって、AI（人工知能）は「人間知能」の4成分の一つである「理性」の一部をまで、代替できる段階まできたと、著者は判断している、ということである。これを図4（前出）に書き入れると、次のようになる。

図5（右ページ）の左下あたりにおいて、「情報」と「知識」の間（「理性」の領域の一部）に、「GPT-3が得意な領域」というキーワードを記入した。

その上でさらに自問自答を続ける。図5の「情報と知能の発展モデル2」において、人間の得意な領域はどこだろうか。それは、この全部である。なぜなら、人間がそれを発見できるところの「情報」の発展モデル自体が、この図の基礎になっているからである。

では図5の「情報と知能の発展モデル2」において、AIの苦手な領域はどこだろうか。それは、この図5から明らかのように、少なくとも2つある。第1に、人間が知識から知恵を発見するために使う知能であるところの「感性」である。第2に、人間が幾多の知恵を総合して統合的理解を獲得するために必要な「知能」であるところの「悟性」の部分である。

なおこの「悟性」という言葉は、一般的に多くの人間には日常用語にはなっておらず、私も（カント哲学以外では）殆ど使わない。確かに意味が分かりにくい言葉である。「悟性」はドイツ語ではVerstand（動詞verstehen）、英語ではunderstandingと翻訳される。一般的には「悟性」は、人間が何かを理解するための全般的な能力を意味し、思考の能力、すなわち人間知能の全体をまとめるもの、というほどの意味を込めて私は使っている。

以上の考察を通して、自然言語処理の最近の発展で知られるGPT-3の位置付けを、人間知能の全体において行なった。

3 おわりに：まとめと今後の展望

この論考の報告で議論したことをまとめる。著者は以前から、そもそも「情報」とはどういうものかを考えることを、大学における一般情報教育の柱あるいは出発点におくべきだと考えてきた。この論考でもそこから出発して、著者が提案する「情報の発展モデル」を紹介した。またAI（人工知能）が日常生活や情報教育に入り込む時代になり、今後の社会に必要な「人間知能」の理解を深めるために著者が提案する「情報と知能の発展モデル」を振り返った。しかし並行して、自然言語処理の分野で劇的な進歩が起こり、それを象徴するGPT-3（2021年現在）を紹介しつつ、AIの言語能力の意味を議論した。その上で自然言語処理のAIあるいは典型的なGPT-3の言語能力を、人間知能の全体モデルである「情報と知能の発展モデル」の中に位置付け、その意味を議論した。

最後に今後の課題について触れる。人間にとって最大の謎は、自分自身である。あるいは人間が「心」と呼ぶ何かである。あるいはこの人間を包み込む宇宙や自然、社会といった、当たり前存在こそが、最大の謎である。古来から人間は人間・社会・自然という不思議な存在に限りない探求の目を向けてきた。この中で情報技術は進歩し続けているのである。特に近年のAI関連技術の進歩は、そもそも人間とは何か、人間の知能とはいかなるものか、という根源的な問いを、再考させる段階にきている。

本論考は、このようなAI技術の意味を再考することで、実は人間とは何かという謎の解明に繋がること、またこの謎の解明こそが今後の課題であると考えられることを指摘したい。また今後の大学での一般情報教育等の教養教育においても、このような根源的な問いの重要性が増していると考えられることを最後に指摘したい。

参考文献

1. 情報処理学会一般情報教育委員会編『一般情報教育』、オーム社、2020年.
2. OpenAI, "Language Models are Few-Shot Learners" (GPT-3), <https://arxiv.org/pdf/2005.14165.pdf>, 2020年.
3. 水野義之「情報社会における「情報」の発展モデル」『日本社会情報学会第24回全国大会研究発表論文集』pp.184-187, 2009年.
4. 水野義之「AI人材の情報倫理教育におけるインフォームド・コンセントを基盤とした能動的学修」『私立大学情報教育協会 教育イノベーション大会』, 2019年.
5. 水野義之「文科系大学におけるICT教育を再考する～アクティブ・ラーニング(AL)から人工知能(AI)の時代へ」、阿部勘一編著『ICT教育再考～文科系大学におけるICT教育の現状と課題』所収、noa出版, 2020年.
6. Y.Mizuno, "Modeling of human intelligence applied to general education of informatics in AI era", AXIES2020, <https://axies.jp/conf/conf2020/>, 2020年
7. 水野義之「AI（人工知能）の理解を目的とする「人間知能」のモデル化提案と情報教育の改善」『現代社会研究科論集』（京都女子大学大学院）, 15号、2020年.
8. D.Jacob, et al., "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding", <https://arxiv.org/abs/1810.04805>, 2019年.
9. A.Radford, et al., "Language Models are Unsupervised Multitask Learners" (GPT-2), <https://d4mucfpksywv.cloudfront.net/better-language-models/language-models.pdf>, 2019年.
10. Lex Fridman, "Deep Learning State of the Art (2020)", <https://www.youtube.com/watch?v=0VH1Lim8gL8>, 2020年.
11. A.Vaswani, "Attention Is All You Need", <https://arxiv.org/abs/1706.03762>, 2017年.