

デジタル人文学の可能性に関する考察 ～大規模言語モデルの登場による学際的分野の革新～

A Study on the Possibilities of Digital Humanities ～Transformation of Interdisciplinary Area through Large Language Models～

大野邦夫[†]
Kunio OHNO[†]

[†]株式会社モナビITコンサルティング [†] Monavis IT Consulting Co. LTD.

Email: k-ohno@star.ocn.ne.jp

1. はじめに

大規模言語モデル (LLM) を用いる生成AIの急速な普及に伴い、人文学研究やリベラルアーツ教育は転機を迎えていると思われる。将来を展望するには過去から学ぶことが必要である。本報告では、過去の人工知能技術、過去の人間の知識に関する哲学的な分析を通じて、今後のデジタル人文学のあり方を考察する。以前の人工知能としては、マービン・ミンスキーによるフレーム理論を基礎とし、推論機構としてプロダクションシステムを用いるエキスパートシステムが構築されたが、ミンスキーのフレームは属性と値のペアを基本とし、データベース技術やオブジェクト指向、マークアップ言語はその枠組みを提供する。哲学分野における人間の知識は、パートランド・ラッセルが晩年に考察しているのもその概要を紹介し、その両者から生成AIの今後の動向、ならびにデジタル人文学の展望について考察する。

2. 生成AIのインパクトと動向

ChatGPTの出現に端を発して、生成AIに関するセンセーショナルな報道がもてはやされているが、事実に基づく論理的、客観的な把握が必要である。2年前の年次大会で、「デジタル人文学とメタデータ」というタイトルで、メタデータの自動生成の可能性について論じた経緯がある。その際には、word2vecやBERTによる言語モデルを紹介し、Open AIのGPTについても紹介した。その際には双方向で分析するBERTに対して単方向でしか処理しないGPTは技術的に劣った技術と考えられた。ただTransformerによるタスクを検討することにより、メタデータ生成は潜在的に可能な状況にあることが推察されたが、もはやそのような状況を通り越して現状のChatGPTは一步も二歩も先に進んだ状況であると思われる。

ChatGPTに関する雑誌の記事として、日経サイエンス5月号[1]、Newton7月号[2]、世界7月号[3][4][5]の特集記事を読んだ。前二者は、技術内容の無難な紹介と、社会的なインパクトに対するチュートリアル的な内容であった。技術的な内容についてはトピック的な取り上げ方が主で、系統だった紹介ではない。岩波の世界は、社会的なインパクトを取り上げていたが、社会を変化させそれが多くの人の幸福に結びつくという視点よりは、人々の自由な思考や意見表明に対する危険性を指摘するもので、社会に警鐘を鳴らす趣旨であると感じられた。

ChatGPTを自分でも使ってみたが、データが豊富に収集された分野の質問では極めて的確な解答をすると感じた。虚偽

回答はデータ蓄積が不十分な分野であり、語彙の出現確率で文章を生成するので正誤チェックが不十分なためと想像された。以前BERTの論文[6]を読んだ際に、単方向では正誤チェックが不十分になるとの指摘があったので、ChatGPTはBERTに比べて劣るのは仕方がないと思われた。BERTに関しては入門的な専門書[7]を読んで、文章のベクトル化、固有表現抽出や文章穴埋め、文章分類、Transformerの役割などについてのLLMの初歩的な内容は把握した。しかしChatGPTを特徴付ける対話タスクに関しては不明である。

そこでChatGPT自身にBERTの双方向性に対するChatGPTの単方向性の適否について質問したところ、BERTに一日の長があることを認める回答であった。自分にとって都合が悪いことを素直に認めるので、却って好感を持てた。次にワイゼンバウムのイライザについて質問してみた。イライザは1960年代末に、精神科医が患者との自然な会話を実現したので話題になったシステムである。患者の話す語彙を引用して話題を拡張するだけの簡単なアルゴリズムの対話システムであるが、患者がその対話からヒントを得て新たな心境になることで知られていた。ChatGPTにイライザについて質問し、的確に回答した後に、その技術を活用しているかと質問したら、それに対しては否定し、深層学習のRNN, LSTM, Transformerを対話に用いていると回答した。確かにその通りであろうが、深層学習もイライザも原理としてはパターン照合なので、強弁かもしれないが同様な側面も存在すると言えるだろう。

以上の様に、ChatGPTはかなり説得力のある回答をすることが可能であるが、その詳細な中身は知ることが困難であった。以上の状況をフェイスブックに投稿したところ、やはりChatGPTを試用して種々の検討をしている方々からコメントを頂いた。例えば、Bingのバージョンを用いると出典を示してくれるとか、Perplexity AIというサイトは類似の質問を紹介してくれるとのことで、生成AIが日進月歩で着実に進展していることが垣間見られるのであった。

3. 従来のAIとの対比

ChatGPTは突如出現したように感じられるが、その背景はグーグルのBERTやELMoのような大規模なコーパスにより文章を分析・生成する言語モデル技術である。GPTは、Generative Pretrained Transformerの略で、事前学習された生成トランスフォーマを意味する。BERTはBidirectional Encoder Representations from Transformersの略で、トランスフォーマによる双方向の符号化表現を意味するが、トランスフォーマは

Huggingface社が提供するオープンソースで、深層学習においてAttentionと呼ばれる注目すべき単語を抽出するメカニズムを具備している。ELMoはEmbeddings from Language Modelsの略で、言語モデルの挿入を意味するが、深層学習におけるLSTM (Long-short Term Memory) を双方向に適用し、単方向に比べ優れた性能を実現した。しかしトランスフォーマ技術

術は使用していないので性能的には不十分であった。BERTは、GPTとELMoの良いところ取りで、トランスフォーマを双方向に用いてGPTを差別化し、優れた機能・性能を実現した[6]。なおこの内容に関しては、2年前の年次大会でBERTの論文に掲載された比較図(図1)と共に紹介した[8]。

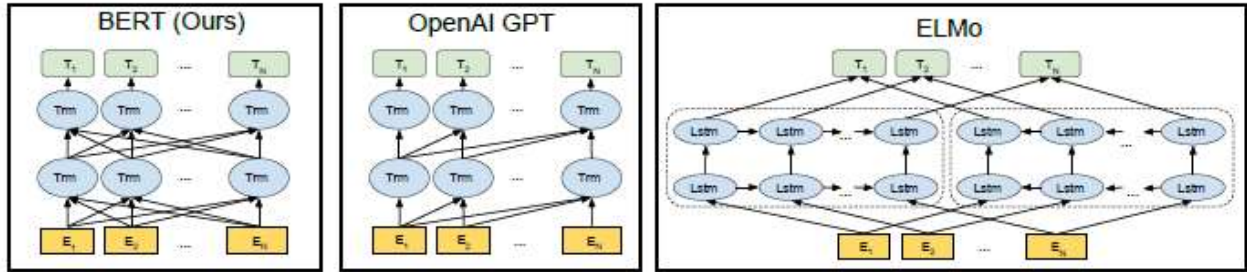


図1 BERTとOpenAI-GPT, ELMoの処理比較

言うまでもなく、深層学習による自然言語処理が、生成AIの基本技術である。深層学習の基礎はニューラルネットで、7~8年前に日本で話題になり、その後は深層学習による画像認識や物体認識が画像電子学会でも研究分野になっている。その当時の議論では、自然言語処理の場合、文章の長さが多様であることから、RNN (Recurrent Neural Network) とLSTMが有効と指摘されていた[9]。別の専門書[10]では、bag-of-wordsモデルやskip-gramモデル、階層型ソフトマックス計算といった基礎的な紹介がなされているが、Googleはそれらの技術を活用するword2vecを既に公開しBERTの開発を、Open AIはGPTの開発を進めていた。その当時、個人的にはAIによって置き換えられる職業に関心を持ち、深層学習について学習・考察し、自然言語処理にこの技術が使われるようになると社会的なインパクトが大きいことを指摘した経緯があるが、それが現実になった観がある[11][12]。

4. エキスパートシステムにおける推論技術

深層学習以前の人工知能技術は、専門家の知識を模擬したエキスパートシステムで、知識ベースと推論エンジンで構成されるシステムであった。知識ベースはミンスキーのフレーム、推論エンジンはif-thenルールによるプロダクションシステムで構成され、LISP言語により実装された。ミンスキーのフレームとプロダクションシステムは、WinstonのLISPの教科書[13]で分かりやすく紹介されていたので、LISPを学ぶ際に関心を持ち習得することができた。

その当時、KEEやARTというエキスパート・シェルと呼ばれる開発ツールが話題になったが、そのセールスポイントはGUIで、知識ベース構築やデモ効果をアピールしたものの、基本はフレームとプロダクションシステムであった。その後、ミネアポリスで開催されたAAAI'88に参加する機会があり、"Exploring Artificial Intelligence"という書籍を入手し、米国のエキスパートシステム技術の先端分野を把握した[14]。その当時最も興味を持ったのは定性的推論 (Qualitative Reasoning) で、事実に基づく普遍的な情報を深い知識と呼び、数理論理的な情報や自然科学的な情報を体系的にミンスキーのフレームで表現するアプローチが検討されていた。この深い知識が後にオントロジー技術としてセマンティックWebに活用されるようになった経緯がある。

"Exploring Artificial Intelligence"には、定性的推論を含め、表1に示すような当時日本では知られていなかった新規の応用推論システムが紹介されていた。

表1 エキスパートシステムにおける応用推論システム

推論手法	内容
説明ベースの推論	一般的概念の説明のパターンをルール記述し、それらの変数の一部を固定して具体化し、それに基づいた特殊概念を用いて推論する仕組み
計画実行推論	LTL (Linear Temporal Logic) という直線的な短期間の因果関係の把握に基づく推論の仕組み
定性的推論	事物の内在的・意味論理的な概念 (オントロジー) を抽出し、その関係性に基づく推論
モデルベース推論	対象を論理的、数理的なモデルで記述し、それに基づき推論する仕組み
エビデンシャル推論	ベイジック推論手法の応用としての Dempster-Shafer (D-S) 理論による条件的な確率に基づき逐次推論する仕組み
テンポラル推論	時間スケールに基づく推論の仕組みで、変化に基づく手法と時間的推移による手法
非単調推論	新たな知識の学習によって既存の知識が覆されるような場合で、デフォルトによる推論、アブダクション、知識に関する推論、信念更新などが挙げられる
インテリジェントチューター	コンピュータが学習者の反応を分析して、なぜその問題を間違えたかを判別しながら状況にあったフィードバックを返す仕組み

これらの推論方式を紹介するのは、最近のLLMの推論タスクとしてヒントを与え得ると感じるからである。その背景としてフレームが自然言語に関係することが挙げられる。Winstonの教科書では、フレームは知識概念でありそれは属性によって構成されるプログラム記述を提供する。属性は値のた

めのスロットを持ち、それを修飾するファセットを持つ。さらに動的な機能としてのデーモンを具備するのであるが、その概念の枠組みはオブジェクト指向のクラスで定義され、具体的な実装はインスタンスで表現される。他方自然言語の文章表現の多くは、構文解析を通じて属性と値の関係、動的にはメソッド名称と処理の関係で記述され得る。LLMとの整合は必ずしも明確ではないが、日時、地名、人名、組織名などを属性として管理する固有表現抽出の延長として今後期待される。

5. ラッセルが想定した人間の知識

バートランド・ラッセルは人間の知識に関して幅広くかつ深く考察した哲学者の一人であろう。元々は数理哲学、論理分析哲学の専門家であり、ホワイトヘッドとの共著によるプリンキピア・マテマティカは代数学を論理学として位置づけたマイルストーンである[15]。

最晩年の哲学書である「人間の知識」[16]は、序論において全宇宙において人間の知識は取るに足らない微小なものであり、それが全世界を定義しているかのような妄想を戒めているが、哲人ならではの慧眼であろう。「人間の知識」は、「科学の世界」「言語」「科学と知覚」「科学的概念」「蓋然性」「科学的推理の要請」の6部から構成されている。この内容から分かる通り人間の知識は科学的知識が基本であり、それは経験的、観察的事実に基づき、推論された仮説を検証するプロセスが重視される。その検証も多くは自然言語の語彙によるものなので、そのあいまい性や蓋然性の問題があり、統計的・数学的な手法が重要になる。そのラッセルが、人間の知識を結論的に、「不確実で不精密で部分的である」と主張しているのである[17]。その背景は、人間の認知機能と自然言語の語彙を背景とする意味概念に起因する。自然言語に起因する問題は、生成AIのLLMにとっても全く同様である。

さらに晩年の「私の哲学の発展」[18]でも11章の「知識」、12章の「意識と経験」、13章の「言語」、14章の「普遍者と個別者と固有名」、15章の「真理の定義」、16章の「非論証的推論」では「人間の知識」を補う観点で人間の知識の不完全さを鋭く指摘している。

16章の非論証的推論は、演繹や帰納といった従来の論理学が扱ってきた推論ではない、言語のあいまいさ、蓋然性に基づく多様な概念を包含する推論を扱う。ラッセルはそれを(1)準永続性、(2)因果線の分離可能性、(3)空間的・時間的連続性、(4)空間時間構造、(5)類比性という5種類のカテゴリに区分して解説しているが、(1)~(3)は主に時空間における近傍の概念に基づく推論、(5)は類似現象(アナロジ)からの類推である。

(4)が構造に基づく推論で、モデル、アーキテクチャ、データ構造といったコンピュータが行う検索や推論機能はこのカテゴリである。データベースはもちろん、ミンスキーのフレームに基づく知識ベースや推論エンジンはこのカテゴリに含まれる。

表1に示したエキスパートシステムの、計画実行推論、定性的推論、モデルベース推論、エビデンシャル推論、テンポラル推論、非単調推論といった手法はこの分野であると思われるが、基本的には(4)のカテゴリに含まれそうである。なお、計画実行推論は、LTL (Linear Temporal Logic) という直線的な短期間の因果関係の把握に基づく推論の仕組みなので、(1)の準永続性と(3)の空間的・時間的連続性に関係しそうである。テンポラル推論は、時間スケールに基づく推論の仕組みで、変化に基づく手法と時間的推移による手法なので、(2)の因果線の分離

可能性に関係しそうである。このように1980年代後半に試みられた推論方式が、ラッセルの分類による非論証推論に対応付けられるのは興味深い。

6. 人文学研究のあり方

人文学分野の研究は、既存の論文や文学作品、雑誌、新聞の記事などの文章の論旨に基づき考察・論証し、新たな知見を主張するものであろう。歴史的には西欧中世に人間を自由に活動させる技芸としての哲学を中心とするリベラルアーツに端を発する。

今日の社会でリベラルアーツは大学の一般教養で、社会人としてのあり方を学ばせる学問分野であるが、産業社会の進展に伴う専門分野の細分化に伴い、総合的な人間性や文化的価値観、さらにその多様性や地球環境における持続性など、従来の人文学が着手していなかった分野の開拓が期待される。私も職業訓練教育に関わるようになり、10年前に異文化コミュニケーション学会(SIETAR Japan)に入会し、SIETAR Europaでも活動したが、広範な背景文化に基づく人間や社会を対象とする課題に対する説得力ある問題解決の困難性を感じている[19][20][21]。特に日本の人文学分野は海外との自由な意見交換を伴う交流が難しいことを痛感する。その背景には、日本文化を支える日本語の問題があり、情報化・IT化によりグローバルネットワーク社会になっても多くの制約が存在する[22]。

日本文化を海外に知ってもらうための情報発信については、デジタルサイネージの活用を含むDSG研究会の課題として以前検討したことがある[23][24]。

そのような観点で、興味深い論文が目についた[25]。この研究は、語彙の相関関係を因果関係として推論する有向グラフを活用する既存論文のアルゴリズムに基づき、LLMを評価するCORR2CAUSEというタスクを構築し、主にGoogleのBERTとOpen AIのGPTに関連するLLMを対象に17システムについてのベンチマーク評価をまとめたものである。論文のConclusionは以下の通りである。

「この研究では、相関関係から因果関係を推測する新しいタスク CORR2CAUSE を導入し、40万サンプルを超える大規模なデータセットを収集した。この新しいタスクに関して広範な LLM を評価し、既製の LLM ではこのタスクのパフォーマンスが低いことを示した。また、微調整することによりこのタスクで LLM を再利用できることも示したが、今後の作業では対象範囲外の一般化問題に注意する必要がある。好意的な思惑を回避するには、このデータセットを使用して、このデータセットを適用していない LLM の純粋な因果推論スキルのベンチマークを行うことをお勧めする。現在の LLM の推論能力が限られていること、および実際の推論と事前学習コーパス由来の知識を分離することが難しいことを考慮すると、我々の研究グループは今後両方の能力を正確に峻別して測定することを目的とした作業が不可欠である。我々は今回の取り組みがその第一歩であると信じている。」

この論文の表4、表5の結果から、GPTに比べるとBERTに一日の長があることが認められる。その背景としては、Transformerを単方向に作用させるよりも双方向に作用させるアルゴリズムの優越性に起因すると推定されるが、論文では言及されていない。実際の推論と事前学習コーパス由来の知識を峻別することが言及されているが、これはラッセルが晩年に

提起した非論証的推論の煩雑な問題をLLMで処理可能な語彙・言語の問題と分離する試みとも感じられ、興味深いではあるが超難題であろう。

7. 考察

以上、2章でLLMを用いる生成AIの急速な普及に伴い人文学研究やリベラルアーツ教育が転機を迎えている状況を述べ、3章で生成AIの技術的背景である深層学習を活用するLLMについて解説し、4章で深層学習以前のAIであったエキスパートシステム技術との関係を整理し、5章でAI以前の人文学の立場で人間の知識を考察したパートランド・ラッセルの業績について紹介した。次に6章で人文学としての現在の立ち位置を考察し、その状況にインパクトを与えそうな最新の研究動向について述べた。

LLMは、世の中に現存する全ての情報を取り込みつつあるので、既存の図書館や博物館、美術館などを含む過去の記述・記録された情報が言語モデルに取り込まれていくであろう。今後多様な言語モデルが開発され、多分野における多様なモデルがサービスに供されるであろう。だが、言語モデルが大規模化するに伴い、その機能・性能は統一化の方向に向かうのではないと思われる。少なくとも過去のデータに関しては、

というのは、自然な対話応答をするためには、常識的な知的背景が必要であり、それが多様化するとは思われないからである。異なる生成AIが同じ問いに対して異なった回答を行なったとすると、その真偽が確かめられ誤った回答をした生成AIの言語モデルは修正を余儀なくされるであろう。

広告代理店や政治権力者は、意図的な情報を客観性を装って回答する言語モデルを作成したいであろう。だがこれは難しいと思われる。その理由は、意図的な結論のための背景情報に真実らしさをもたらすために、ニューラルネットワークの特定ノード群の閾値を変更すると、意図せざる問題に対する回答に誤りを生じて全体の整合が取れなくなるためである。あるウソを通すために別のウソを付かねばならなくなるような状況である。LLMに取って、“Honesty is the best way.”なのではあるまいか。

同様な議論が、オントロジ分野でも存在した。ドメイン・オントロジは定義可能でも、全分野を網羅するアッパー・オントロジに関する議論が問題であった。自然言語の語彙をOWLのクラスや属性にしてアッパー・オントロジを定義するという発想であったが、これはゲーデルの不完全性定理に抵触するので本質的に無理である。他方、LLMは、アッパー・オントロジに相当する領域に対応する。これは裏を返すと、語彙関係を論理的に定義することは不可能だが確率統計的に関係付けることは可能なことを意味する。CORR2CAUSEのタスク[25]の将来テーマかもしれないが、興味深い課題である。

8. おわりに

以上、生成AIの登場に基づくデジタル人文学の展望と課題について述べたが、将に前途洋々の状況であり、関心を抱く若手の研究者に挑戦して頂くことを期待したい。

なお、本報告を執筆する上で、高度技術者育成と技能伝承研究会（高承研）、並びにラッセル読書会における議論が参考になったので関係者に感謝します。さらに、私のフェイスブックへの投稿に対する人々の反応にも有益で勇気付けられた。

特にKDDIの高木悟様、前NTT研究所の水川 真様、前東京工業大学教授の小野京右様に御礼申し上げます。

文献

- [1] 吉田和輝, 今泉允聰; “AIに人間らしさをもたらした大規模言語モデル”, 日経サイエンス 2023 05, pp.32–37, Mar, 2023
- [2] 松尾豊; “劇的に進化した「対話AI」 ChatGPTの衝撃”, Newton 2023 7, pp.12–41, May, 2023
- [3] ナオミ・クライン (中村俊太郎訳); “「幻覚を見ている」のはAIの機械ではなくその製作者たち”, 世界 2023.7, pp.84–95, Jun, 2023
- [4] 山崎憲; “AIがもたらす<働き方>にどう立ち向かうか”, 世界 2023.7, pp.96–105, Jun, 2023
- [5] 久木田水生: “今そこにある倫理的問題～AIをめぐるリスクと規制”, 世界 2023.7, pp.106–105, Jun, 2023
- [6] Jacob Devlin, Ming-Wei Chang, Kenton Lee, Kristina Toutanova; “BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding”, arXiv.org > cs > arXiv:1810.04805, Cornell University, 2018
- [7] 近江崇宏, 金田健太郎, 森長誠, 江間見亜利; “BERTによる自然言語処理入門”, ストックマーク株式会社編 オーム社, 2021
- [8] 大野邦夫; “デジタル人文学とメタデータ”, 2021年度画像電子学会年次大会講演論文, Jun, 2021
- [9] 岡谷貴之; “深層学習”, 講談社, pp. 111–130, 2015
- [10] 神高敏弘, 麻生英樹, 前田新一, 岡野原大輔, 岡谷貴之, 久保陽太郎, ボレカラダヌシカ; “深層学習”, 人工知能学会監修, 近代科学社, pp. 211–247, 2015
- [11] 大野邦夫; “深層学習技術の適用に関する一考察～第2世代AI事業関係者の個人的な見解”, 第3回画像関連学会連合大会講演論文, Nov, 2016
- [12] 大野邦夫; “深層学習時代のIoTサービスと人材育成への展望と課題”, 画像電子学会第7回DSGワークショップ, Nov, 2016
- [13] P.H.ウインストン (白井訳); “LISP”, 培風館, 1982
- [14] Howard E. Shrobe et. al., “Exploring Artificial Intelligence: Survey Talks from the National Conferences on Artificial Intelligence”, Morgan Kaufmann Publisher, Inc., 1988
- [15] A・N・ホワイトヘッド, B・ラッセル (岡本賢吾, 戸田山和久, 加地大介訳); “プリンキピア・マテマティカ序論”, 哲学書房, 1988
- [16] バートランド・ラッセル (鎮目恭夫訳); “人間の知識・II”, バートランド・ラッセル著作集 No.9・10, みすず書房, 1970
- [17] バートランド・ラッセル (鎮目恭夫訳); “人間の知識II”, バートランド・ラッセル著作集 No.10, みすず書房, p.412, 1970
- [18] バートランド・ラッセル (野田又夫訳); “私の哲学の発展”, みすず書房, 1960
- [19] 大野邦夫, 西口美津子, 渡部美紀子, 末永早夏; “異文化交流スキルを有する女性起業家の育成に関する研究”, 2014年度異文化コミュニケーション学会年次大会研究報告 Sep, 2014
- [20] Kazunori Akutagawa, Kunio Ohno, Mitsuko Nishiguchi, “Human Resource Development of Woman Entrepreneurs in Fukushima with Intercultural Historical View”, Proc SIETAR Europa 2015 Congress, May, 2015
- [21] 大野邦夫, 渡部美紀子, 西口美津子, 阿部鳴美, 吉田美意子; “スペインtailを通じた女川とスペインの文化交流”, 2017年度異文化コミュニケーション学会年次大会研究報告 Oct, 2017
- [22] 大野邦夫; “コンテンツ管理における日本語処理に関する考察”画像電子学会第3回デジタルミュージアム・人文学研究会資料 Mar, 2022
- [23] Kunio Ohno, Mitsuko Nishiguchi, Toshiko Kimura; “A STUDY ON DIGITAL SIGNAGE SYSTEMS FOR REGIONAL INFORMATION SHARING”, Proc. 6th International Workshop on Image Electronics and Visual Computing, Aug, 2019
- [24] 大野邦夫; “DSG研究会の足跡と今後の発展”, 画像電子学会誌, Vol.50, No.3, pp431–438, 2021
- [25] Zhijing Jin, Jiarui Liu, Zhiheng Lyu, Spencer Poff, Mrinmaya Sachan, Rada Mihalcea, Mona Diab, Bernhard Scholkopf; “Can Large Language Models Infer Causation from Correlation?”, arXiv:2306.05836v1 [cs.CL] 9 Jun, 2023