科学・工学と人工知能技術の融合を目指して

Towards Integration of Science and Technology with AI Systems

国立研究開発法人産業技術総合研究所　フェロー　人工知能研究センター　研究センター長

辻井　潤一

国立研究開発法人産業技術総合研究所　フェロー、人工知能研究センター　研究センター長、英国マンチェスター大学教授、国際計算言語委員会（ICCL）委員長、AAMT ／ Japio 特許翻訳研究会委員長

1　はじめに

第３期の人工知能ブームが喧伝され出してから、すでに 10 年あまりの時間が過ぎた。この時間の経過に伴って、人工知能技術の影響の大きさとその限界も次第に認識されてきた。人間生活の様々な局面に浸透し、いままで想像できなかった変革を社会にもたらす可能性があることが認識され、その過程で、技術の持つ肯定的な側面だけでなく、否定的な側面も意識されるようになってきた。

人工知能技術を我々の日常生活の外側にあるものと考えてきた時代から、日常生活の内部に浸透しそれを大きく変革する技術であること、このことを多くの人が自分事として考えるようになった、ということであろう。スタンフォード大学、ケンブリッジ大学など、有力な研究機関が Human Centric AI（人間中心の AI）を標榜し、技術研究としての AI 研究から、人間科学や社会科学との連携を重視した組織を立ちあげて、人間知能と人工知能との相互関係を対象とする研究を始めている。

私が属している産業技術総合研究所・人工知能研究センターは組織の性格から、人間科学や社会科学と連携していくというよりも、技術の観点から、この問題に取り組んでいる。人工知能が単独で知的な判断やタスクをこなすという枠組みから、人間知能と人工知能がより緊密に連携することによって、それぞれ単独ではできない知的なタスクをこなしていく枠組み、その技術的な基盤を作ることを目指している。

2　２つの知能

我々は知能というものを一つの能力と考える傾向がある。その能力の高さや低さは一次元的に並べることができて、人 A は人 B に比べて知能が高いとか、碁をさす人工知能が人間のプロに勝ったことから、人工知能が人間の知能よりも高くなった、と考える。多くの人はこれが単純化であることはわかっているが、なんとなく知能という一つの能力があり、優劣の比較ができると思ってしまう。

このことが単純化であることは、知的な能力を必要とするあらゆる仕事で、常に人 A が人 B に勝る、ということが稀であることからも、明らかであろう。人の知能を一次元に並べて数値化する知能指数も、実際には、空間把握の能力、言語能力、数字の操作能力といったように、知能を構成する要素的な能力に分けて測定し、便宜的にそれらの結果を一次元的な数値に置き換えているに過ぎない。

人工知能と人間知能のように、「知的な能力」を支える基盤そのものが異なる知能の場合には、このことは顕著であり、「知能」という一つの能力やそれを測る尺度が両者に共通にある、と考えることはできない。

第３期の人工知能は、データの統計的な解析やデータからの規則発見を目指した Big Data の時代を経て、その延長としての機械学習を中核技術として発展してきた。

例えば、医療診断においても、過去の数百万件の患者さんの検査データがあれば、新たな患者さんの検査デー

タから、その患者さんの診断をしたり、確率的にみて最適の治療方法を提示したりすることができる。診断能力（すなわち、判断能力）は、知的能力の重要な一つである。膨大な患者データから学習することで、計算機がその知的な能力を獲得したことになる。

経験を積んだ人間の医師も、多くの患者さんの検査データや時間的な経過を観察することで、経験の浅い医師よりも優れた診断能力を獲得する。そういう意味では、人間の判断能力の獲得にも、人工知能が判断能力を獲得する学習過程と類似の過程がある。人間の判断能力が、観察データからの帰納だけで獲得されたものであれば、数百万の検査データから学習する人工知能に負ける。人間には、大規模なデータから、そこに潜む潜在的な規則性を一般化してとらえる能力はない。

ただ、少し冷静に考えてみると、検査データとしてどのようなデータを採取することが有効か、あるいは、治療を行う上で病疾患をどのようなカテゴリーに分けて考えるべきかなど、検査データと診断とを結びつける枠組みは、医療科学の長い歴史の過程で医療科学者が作りだしてきたものである。

表面的に同じような症状があることから同一の病疾患カテゴリーと考えられてきたものが、実は、その背後にある機構が違うことから別の疾患カテゴリーに分割され、それぞれを区別するための検査データの取得方法や治療法が作り出されてきた。この過程が医療科学の発展であり、その結果として、現在の検査データと診断結果の対のデータが集積されている。こういう枠組みがなければ、現在の人工知能技術であるデータからの病理診断という枠組みそれ自体が成立しない。

以下では、人間と人工知能の持つ「知能」の差を、科学・工学と技術との関係から、考えてみることにする。

3 科学とエンジニアリング：end-to-end の人工知能技術

人工知能技術と総称されるものは、社会実装が進展するに伴って、いわゆる深層学習以外のさまざまな情報処理技術と組み合わされたり、第 1 期や第 2 期の人工知能技術が実装されたりと、その多様化が進展している。ただ、現在、人工知能技術がこれほど喧伝され、社会に大きな影響を及ぼすようになったのは、Big Data、機械学習、深層学習の流れ、いわゆるデータ指向型人工知能（Data-Oriented AI）の隆盛があったから、である。

以下の論考では、人工知能技術というときには、特に断らない限り、このタイプの人工知能技術を指すことにする。また、人工知能を AI と略称する。

さて、この Data-Oriented な AI 技術を象徴する言葉に、end-to-end がある。図 1 に示すように、入力の観察データと正解となる出力のデータの対を大量に与えると、その 2 つのデータ間に存する計算関係を AI システムが学習する、という枠組みである。例えば、医療診断のデータを検査データと診断結果の病名とに分けて、前者を入力データ、後者を出力データとしてシステムに与えると、その間に成り立つ規則性（計算関係）の一般化を AI システムが学習する。

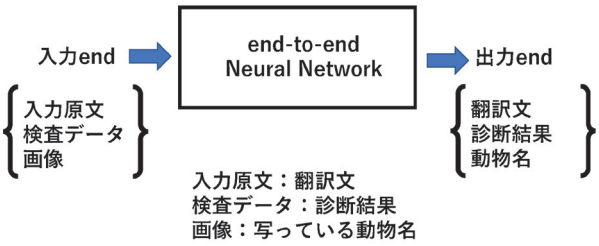

図 1　end-to-end システム

あるいは、動物が写っている写真（画像データ）とその動物の名前の対を、入力・出力の対として与えると、新たに与えられた写真に写っている動物名を出力する AI システムができる。

英語の文とその翻訳の日本語文を入力・出力の対として大量に与えると、英語・日本語の翻訳関係にある文の計算関係を AI システムが学習し、英語・日本語の翻訳システムが出来上がる。

この end-to-end の枠組みは、大量の入力・出力の対を集めれば、入力という end と出力という end を結びつける一般化された計算規則を AI システムが学習できる、というわけであるから、非常に強力である。2 つの end を結びつける規則を人間が解明する必要はない。

第 3 期 AI 技術の以前には、例えば、猫と犬の写真を区別するための規則を人間が捉えて、それをプログラムとして実現しようとしていた。猫と犬を区別する規則を明示的なアルゴリズムや規則に書き下せと言われても、そういう規則を書き下すことは容易ではないし、不可能であろう。

機械翻訳の場合も、同じである。end-to-end の翻訳

システム以前は、原文の文法構造と翻訳文の文法構造との対応規則、あるいは、２つの言語における単語や慣用句の対応規則といったものを人間が考えて規則を書き下すことにより、翻訳システムを開発しようとしていた（図2）。ところが、実際にこういう規則を書き下そうとすると、翻訳には、文の文法構造や単語・慣用句の対応といったもの以外にも、文の意味や文脈、話者や聞き手が持っている知識など、多様な要因が複合的に関係していて、全体として、非常に複雑な規則になってしまう。

言い換えると、翻訳の過程でどのような要因が相互に関与し、入力の原文から出力の翻訳文が作り出されるかの過程を人間が明確に把握する。次に、それを規則としてシステムに与えることで機械翻訳を実現するというやり方では、広く実用に耐えるシステムは構築できなかった。

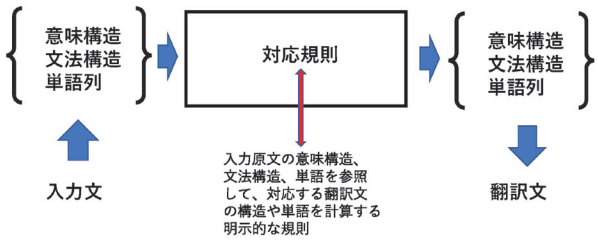

図２　規則による翻訳システム

翻訳の過程を科学的に解明しようとする、いわば翻訳の科学というものと、翻訳システムを構築するというエンジニアリングという２つの試みに実は大きなギャップがあった。現在のend-to-endの機械翻訳システムは、翻訳過程の解明という科学的な試みから、機械翻訳システムの構築を切り離し、大規模な過去の翻訳対のデータだけから翻訳の計算関係をAIシステムが自己組織的に学習している。

このような科学とエンジニアリングの乖離[(1)]は、現在のAIシステムが成功している分野に共通してみることができる。写真に写っている動物を区別するための規則、検査データから病名を判定する診断システムも、それぞれの規則を書き出すことが困難な問題であった。人間が対象を理解する試み（いわば、科学的な思考）をエンジニアリングに結びついけることができなかった問題に対して、end-to-endが強力な解決手段を提供してきたことになる。

4 科学と切り離されたエンジニアリングの限界

よくわからない問題領域でも、入力データと出力データの対を大量に与えれば、問題を解くための一般的な計算関係をAIシステムが学習するというend-to-endの手法は非常に強力であり、多くの問題領域に適用され成功している。ただ、この枠組みの限界も、次第に意識されるようになってきた。

現在のAI技術の欠陥とされるシステムのBlack Box性は、その典型である。データから自己組織的に学習された結果が、外部の観測者（人間）には理解できない。

学習された結果が、実は、与えられたデータが持つ偶然的な性質に左右されていて、新たなデータに対する判断がその偶然的な性質に左右され誤った判断をする可能性が排除できない。あるいは、判断の過程のBlack Box性のために、判断根拠が示されず、天下り的な判断のみが出力される。このような天下り的なシステムが使えないのは、医療診断への応用など、多くの応用にみられる。

現在のend-to-endの機械翻訳システムは、自然な翻訳文を作り出すことができる。ただ、自然に読めるが、原文にない情報が翻訳文につけ加えられたり、逆に、原文にある情報が落とされたりする。翻訳文は自然に読めるが、誤訳になっていることもある。

さらに、このような誤りが発見されても、その誤りが出ないようにシステムを改編することも容易ではない。このシステムの改編の困難さも、システムが内部でとらえている計算関係がBlack Boxであることから生じている。

翻訳のある部分が、原文のどの部分に対応しているのかがわからないTraceabilityの欠如は、機械翻訳だけの問題ではない。画像による医療診断において、病名や病気の進行度合いを天下り的に示されても、医師はその判断を採用することはできない。画像のどの部分のどういう特徴をもとにその診断がなされたのかのTraceabilityがないと、判断がむつかしいケースの判断や自分の判断と異なる判断をしたAIシステムの結果を医師は信頼して採用することはできない。Traceabilityの欠如は、今のAI技術のBlack Box性の表れの一つである。

すこし視点が異なるが、end-to-endシステムの深刻

な問題に、観察データの状況依存性がある。観察データは、そのデータが生じる背後の機構の結果として、観察可能なデータとなる。この背後の機構が安定し、不変である場合には、過去のデータで学習された end-to-end の計算関係は有効である。

ただ、多くの応用分野において、この背後にある機構そのものが変化する。背後にある機構の変化によって過去の蓄積データが有効性を失う。このわかりやすい例に、コロナの感染状況やピークアウトを予測する上での過去のデータの有効性の問題がある。コロナ感染が拡散した初期のころから、ウィルスの mutation によって、次々に変種のウィルスが現れ、過去のデータに基づく予測モデルは役に立たなくなった。

このようなデータの状況依存性に対処するためには、観測されるデータの発生機構そのものへの理解が必要であり、end-to-end でデータだけから構築されたシステムの適用範囲の限界となる。背後にある発生機構の変化に対応するためには、発生機構そのものに関する理解が不可欠である。

end-to-end という極めてエンジニアリング色の強い枠組みが、より有効に使われていくためには、データが生成される背後の機構の科学的な理解と、end-to-end で学習された計算関係との相互の関係を明らかにする White-Box 化が不可欠となる。

5 Professionals-in-the-loop

昨年度の YEAR BOOK への寄稿では、Professionals in the Loop の考え方が日本という成熟社会での AI 技術の発展には有効であろうことを論じた。

end-to-end のシステムを対象の理解に結びつけるためには、対象の理解のための体系を作っている専門家（科学者や工学者）と AI 研究者が密接に連携していくことが不可欠である。成熟社会の日本は、それぞれの対象分野での科学者や工学者に恵まれている。この環境を積極的に活用していくことで、日本独自の AI 技術を構築し、それぞれの分野での日本の競争力を高めていくことができるのではないか、という主張であった。

例えば、医療画像を診断に使う AI 技術を考えてみよう。end-to-end のシステムでは、画像を一方の end（入力）とし、診断結果をもう一方の end（出力）として、その間の計算関係を Black Box の深層学習モデルがとらえる。

一方で、最近の深層学習の研究でよくつかわれる注視機構（Attention Mechanism）が判断過程でどのように寄与したかを可視化することができる。すなわち、AI システムが画像のどの部分に注目してその判断を下したかを医師に見せることができる。これにより、完全な Black Box であったシステムの内部的な動きを人間側が把握でき、部分的ではあるが透明性を向上する。実際、可視化された注視の個所を精査することで、AI システムが判断に使った個所が、人間の専門家から見ると意味のない個所であったケースも確認されている。

このような White Box 化によって、医師は、AI システムが無意味な特徴に反応しているのではないことを確認できる。

まだ、完全に成功しているわけではないが、同様な試みは機械翻訳でも試みられている。深層学習モデルの内部での注視機構の動作を観察することにより、入力文のどの個所を見て翻訳文のどの個所が作り出されてきたかの相互関係が把握できる。これにより、入力文中に対応表現がないものが出力文に現れたり、その逆が生じたりする場合を検出することができよう。

6 科学・工学とエンジニアリングとの融合を目指して

合理的・科学的な把握が困難な対象分野、これまで科学的な研究の対象になりにくかった分野において、判断や予測を行うシステムを構築するエンジニアリングの方法論として、データに基づく end-to-end は強力なものであった。

例えば、経験のある医師の診断過程、翻訳家が行う翻訳の過程、ベテランのエンジニアや技術者が経験の中で習得する Know-How などは、科学的な解明や言語化が困難であり、これらの能力を人工のシステムに移行することが困難であった。これに対して、end-to-end のシステムでは、観察データだけから、隠れた規則性をシステムが自己組織的に学習し、人間を代行する AI システムが構築できる。

ただ、多くの分野は、観測可能なデータを生み出している機構が完全に未解明であるというわけではない。例えば、医療分野は医療科学の進展により、病疾患の発現

過程の部分的な理解が深まり、それによって診断の方法も進展している。多くの分野において、完全な予測性は獲得していないが、背後にある機構の部分的な科学的な解明が行われている。

現在の AI 技術は、入力と出力の計算関係をデータから明らかにする帰納的な方法論の典型的なものである。帰納的方法論は、外部に発現するデータだけから規則性を構築するために、そのデータを発現する背後機構の変化に対しては脆弱である。

今後は、この帰納的な方法論に加えて、データの発現機構に関する科学的な知見に基づく演繹的な方法論を融合していくことが必要である。

帰納と演繹を融合する多様な方法論が、今後、開発されていくであろうが、次節では、その具体例として、我々のセンターの研究を紹介する。この方法論は、対象に関する科学的な知見を深層学習のネットワークの設計に反映する試みとなっている。

7 科学と end-to-end の融合

対象とした問題は、有機化合物の構造式からその物質のもつ物性的な特性パラメータを予測する問題で、材料設計の分野での重要な技術となる。科学的な知見だけでは、与えられた構造式から特性パラメータを予測することはできず、これまで蓄積されてきた化合物とその特性パラメートのデータとを使った帰納的なアプローチが必要となる。図 3 は、この研究で使われた 2 つのニューラルネットワークを示している [(2)]。

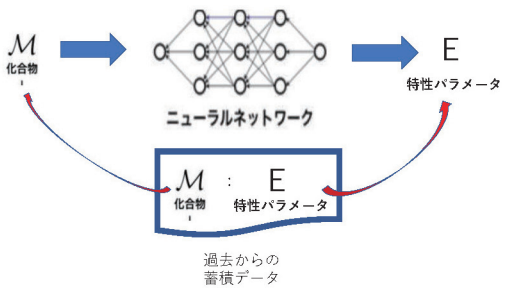

図 3（a） 単純な end-to-end のニューラルネットワーク

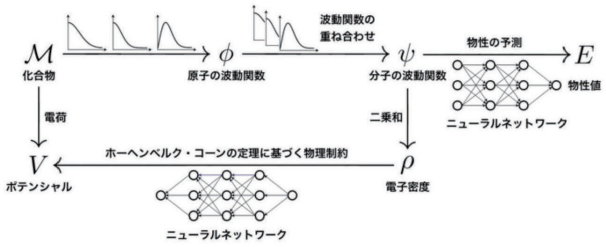

図 3（b） 科学的知見を反映したニューラルネットワーク

実際、有機化合物の構造式とその特性パラメータは、これまでの研究によって、すでに多くの化合物で計測され、データベース化されている。そこでこの予測問題をとくための素朴な手法は、構造式を入力の end、特性パラメータを出力の end とする end-to-end の深層学習システムを構築することである（図 3（a））。

低分子の有機化合物に関しては、特性パラメータの計測が進んでいるために大量のデータがあり、この end-to-end システムはなかなか良い予測精度を示す。しかしながら、予測モデルが実用的な価値を持つのは、高分子化合物に対する予測である。

高分子の構造式はそれを構成する要素とその組み合わせが爆発的に増えるために、特性パラメータが計測されている化合物は、その可能な空間中のごくわずかしかカバーしていない。

そこで問題は、豊富に存在する低分子の有機化合物の計測データで学習されたニューラルネットワークが高分子化合物のパラメータ予測にどれほど有効かであるが、図 3（a）のモデルでの結果は、満足のいくものではなかった。低分子での高い予測精度は、高分子化合物に関しては保持できず、予測誤差は非常に大きくなっている（図 4 の青線）。

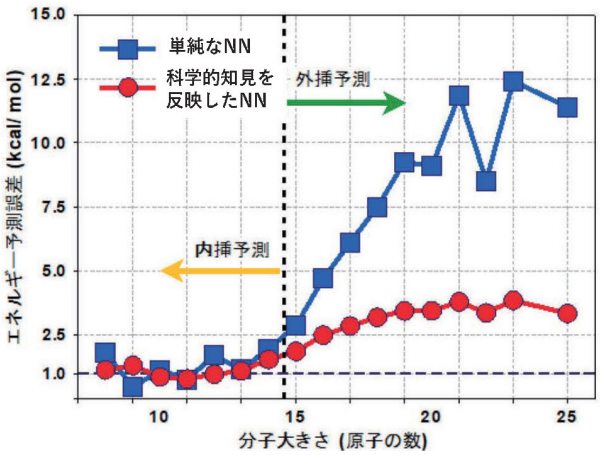

図 4　2 つのネットワークの予測誤差

この結果は、データに基づく帰納的な方法論だけでは、学習データがよく被覆している領域に関しては優れた予測を行うが、被覆がスパースになる領域では予測精度が極端に低下することを示している。

この実験結果は、データからの帰納は学習データが豊

富な領域での内挿（interpolation）による予測はできるが、その学習された計算関係の外挿（extrapolation）の能力は極めて低いことを示している。

これに対して、問題の背後にある一般的な規則性を明らかにしようとする科学の演繹的な方法論は、異なった環境下でも成り立つ一般的な規則性をとらえることで、より優れた外挿的な能力をもつ。

この科学の知見をニューラルネットワークの設計に反映したものが図３（b）である。詳細の説明は省くが、このネットワークでは化合物の構造式と特性パラメータとの間に、この２つの橋渡しをする中間的な表現があり、図３（a）の大きなネットワークをこの中間表現で￥を使って分割している。

この中間的な表現の存在は、量子化学の理論により設定されているものであり、ニューラルネットワークは、構造式とこの中間表現、そして、中間表現と特性パラメータという、２つの具体的な関係を低分子化合物のデータから学習する。

この量子化学の知見をネットワーク設計に反映したシステムは、図４の赤線に示すように高分子化合物に対しても優れた予測精度を持ち、すぐれた外挿の能力を獲得していることがわかる。現在、センターでは、この研究を実際に使える技術にしていくために、物性研究者とAI研究者とのより緊密な連携研究を推進している。

現在のAI研究は、AI研究者だけで研究開発する段階から対象分野の科学者・工学書・技術者と緊密に共同する段階に移行しつつある。

8 おわりに

本稿では、現在のAI技術の中核であるend-to-endのシステム構築に関して、その利点と限界を論じた。これまでの技術は、我々人間が明示的に把握できないことをエンジニアリングの対象とすることには強い限界があった。大量の観察データからの学習を中核とするAI技術は、この限界を破ることにより、人間が明確に把握できていない対象もエンジニアリングの対象にすることを可能にした。

AI技術のこの大きな可能性をうまく生かし、かつ、広い範囲での問題に適用していくためには、データからのAI技術に人間による対象の把握、科学的な知見をうまく融合させていくことが必須となる。

それぞれの分野での優秀な科学者・工学者・技術者がいる日本には、その強さを活かすことで、独創的なAI技術を生み出す可能性がある。大いに期待したい。

参考文献

(1) Junichi Tsujii; Natural Language Processing and Computational Linguistics. Computational Linguistics 2021; 47 (4) : 707-727

(2) Masashi Tsubaki and Teruyasu Mizoguchi: Quantum deep field: data-driven wave function, electron density generation, and energy prediction and extrapolation with machine learning. Physical Review Letters, 2020: 125, 206401