

人工知能の哲学：知能の理解と実現に挑む工学と哲学の対話

知能を人工的に実現する試みである人工知能の研究には、一階述語論理やその亜種に基づく推論の形式化による論理的（記号的）アプローチと、神経回路網のシミュレーションなどによる統計的アプローチがある。人工知能の研究の黎明期において既にこの二つのアプローチは出現していたが、しばらくの間は論理的アプローチが主流であった。この論理的アプローチからは幾つもの重要な研究成果が得られているが、人工知能という名にふさわしい計算機を具体的に作り出すまでには至らなかった。

計算機の性能の向上によって深層学習の技術が発達し、論理的アプローチに代わって統計的アプローチが主流となって著しい成果を挙げたことで、人工知能が広く社会から関心を持たれるようになった。人工知能に関わる技術の爆発的な発展を目の当たりにして、人工知能に関わる様々問題、特に、一台の人工知能の能力が全人類の知的能力を凌駕するシンギュラリティと名付けられた「時」は果たして本当に訪れるのか、訪れるとすれば何時、何が起きるのかといった問題が様々な場所で盛んに論じられるようになった。この種の問題は「技術の発展は科学や社会にどのような影響を与えるのか」という平凡な問題の一種であるが、哲学的にも倫理的にも重要なものであることに疑問の余地はない。

しかし、人工知能はしばしば単なる高性能計算機と同一視され、人工知能に関わる多くの議論の中で、生身の人間には判断不可能な複雑な意思決定問題のような「知能の代替物としての人工知能に固有の問題」と、自動車の自動運転の誤動作のような「科学技術の発展一般に関する問題」が混同されている。この混同は人工知能について考えることを難しくする要因の一つであり、人工知能について意味のある議論を展開するためには、シンギュラリティについての話を始める前に、まず、この二種類の問題の区別を明らかにしなければならず、そのためには「そもそも知能とは何か、知能は人工的に実現可能なものなのか」という人工知能に関わる古風な哲学的問題と向かい合う必要がある。

このシンポジウムでは「哲学と工学・論理と統計」という二つの二項対立を軸に、工学者と哲学者がそれぞれ人工知能に関わる最近の研究を紹介することで、知能の理解と実現における論理的アプローチと統計的アプローチは排他的なものなのか相補的なものなのか、人工知能の技術に関わる工学的な発展は我々の知能の哲学的な理解にどのような変化をもたらすのか、そもそも知能の理解と実現のためには何を明らかにする必要があるのか、といった問題について考えてみたい。

オーガナイザ・司会：菊池誠（神戸大学）

提題者

- 松原崇（神戸大学）「深層学習は何をどのように“学習”するのか」
- 松崎拓也（名古屋大学）「数学における自然言語解析」
- 松王政浩（北海道大学）「確率統計をめぐる人工知能の哲学的問題と哲学者」
- 松阪陽一（首都大学東京）「規則とパターン：後期ウィトゲンシュタインの洞察」

深層学習は何をどのように“学習”するのか

松原崇 (神戸大学)

情報通信技術の発達に伴いあらゆる社会活動が電子化しており、人間が生み出す電子データは指数関数的に増加している。これを受けて現在、データ駆動型という考え方が脚光を浴びている。これは膨大なデータの中から対象の背後にある構造や法則を自動的に“学習”し、解析する手法である。従来のような対象に関する多くの事前知識を用いる解析手法とは、大きく異なる考え方であるとされる。

データ駆動型の代表的な手法の一つが深層学習 (Deep Learning) である。深層学習の多くの適用先のうち画像識別や物体認識では、ただ無数の動画を与えるだけで、自動的に「猫の顔」や「人の体」という抽象的な概念を“学習”したと言われている。そのほか、音声認識、音声合成、自然言語処理、機械制御においても、専門家が設計した緻密なルールやモデルよりもはるかに良い結果を残している。その最たるものとして、今年の3月にGoogle社の人工知能 (AI) AlphaGo が、囲碁の元世界王者に勝利したことが挙げられる。囲碁は一般的な盤上ゲームの中で最も複雑であると考えられており、AI は近年まで強いアマチュアに勝つことも難しかった。AlphaGo は深層学習を用いて、膨大な棋譜データとAI 同士の対局経験から、囲碁における高度な戦略を“学習”した。その打ち筋は開発者のみならず第一級のプロ棋士にも予測不能なほどであった。

もし深層学習があらゆるものを自動的に“学習”してくれるというのなら、我々は解析対象に対して無知なままでも、データという供物をコンピュータという神に捧げるだけで、解析結果というご託宣が得られるようになるのだろうか？私はいくつかの理由により、そのような日が（少なくとも当面は）来ないと考える。大きな理由として、深層学習の成功には、確固とした理論的背景のあるモデルとの組み合わせや、人間の持つ事前知識の支えが不可欠であったことが挙げられる。AlphaGo も強化学習を用いていた。本発表では私が日々四苦八苦している深層学習の「訓練」経験を基に、何をどのように深層学習が“学習する”のか（あるいは、どのようにして深層学習を「訓練」するのか）について述べる。

数学における自然言語解析

松崎拓也 (名古屋大学)

2012年から現在まで、大学入試数学の問題を自動的に解くシステムの開発を続けている。このシステムは日本語で書かれた問題を論理式へと変換した後、数式処理を中心とした自動演繹を行うことで解答を得る。シンポジウムでは、システムの前半部分、すなわちテキストを論理式へと変換する部分を開発する過程で経験したことについてお話ししたい。以下では話題のうち2つについてまとめる。

1. 工学的な課題としての言語理解

システム前半部分の中心は、統語構造に従って各文に対応する意味表現（論理式の断片）を合成する処理である。この処理は組合せ範疇文法という枠組みに従って記述された文法と辞書を基にしている。「枠組みに従って」とはいうものの、文法および辞書の構成方法には相当な自由度がある。また、言語表現と意味との対応における規則性を過不足なく捉え、かつ実際に文解析器が動作するように文法・辞書を構成するための汎用の処方箋は存在しない。動作する（＝予測を提出できる）形で種々の言語現象に関する分析を蓄積した文法と辞書は、物理学における「理論」の意味で日本語（のある部分）という現象に対する「理論」の役割を果たすものである。しかし、新たな現象に出くわす度に文法・辞書は拡張され、その影響はしばしば文法・辞書の全体に渡る大きな変更を伴う。このため、従来の科学における「理論」とは更新の頻度と程度がかなり異なる。このように、言語理解部の中心である文法・辞書は、科学というよりは工学的な成果物の色彩が強い。しかし、言語の表層と意味とが結びつく有様を、科学的検証が可能な形で、かつ、数値パラメータのかたまりとしてではなく言語化可能な形で提示するために、現在ほかに代替手段があるとは思われない。

2. 数学問題テキストにおける時間表現の類似物とその意味

点の運動に関する問題を除くと、大学入試数学で時間の概念が陽に現れることは珍しい。しかし、「～するとき」「～のときの～」「～して（その後）～する」といった表現は、時間に関する表現を借用して「ある状況がある時点で成立していると仮想する（と）」という条件を表しているように見える。これらの条件表現を論理結合子「ならば」を用いて直訳することはしばしば簡単でない。そもそも「～ならば～」の形で簡単に述べにくい条件だから時間表現を借りて表現しているのだと思えば、そのような直訳が難しいことは不思議ではないが、かといって、時間の概念を問題の論理表示に持ち込むことで翻訳が簡単になるようにも見えない。これら直訳が難しい表現の特徴は、「日本語表現の統語構造と（述語）論理での表現の統語構造が対応していない」とまとめることもできる。これが顕著な別の例としては、存在量化を表す「～をうまく選ばば～となる」のような表現もある。これらは「メタファー」つまり直接の意味を再度解釈することで本来の意味が現れる、といった現象なのだろうか。数学テキストでさえメタファーの概念を持ち込まなければ解析できないのだとしたら、そもそも「直接の意味」と「メタファー」を切り分けなければ適用範囲を確定できないような意味の理論は現実の言語という現象に対してはかなり脆弱なのではないか。

確率統計をめぐる人工知能の哲学的問題と哲学者

松王政浩（北海道大学）

私がこのシンポジウムで依頼いただいたのは、確率統計の面で、人工知能に関わる何か哲学的な議論を、ということだ。「確率」「人工知能」「哲学」の三語が並べば、哲学者なら J. Pearl の DAG による統計的因果推論、およびこれに基づく Spirtes らの哲学的議論とそれに対する批判的な議論 (Cartwright ら) の応酬がすぐに思い浮かぶ。確かにこうした議論は「人工知能の哲学」とも言えるが、哲学者の多くはむしろこれを伝統的な因果論争に対して立てられた新機軸と捉えたように思う。実際、ベイジアン・ネットワーク (BN) は確率変数間の定性的関係をグラフで表し、変数間の定量的関係 (依存関係) を条件付き確率で表した確率モデルで、因果推論以外にも様々な確率推論に用いることができるが、哲学者の関心はほぼ因果の問題に集中し、BN の他の潜在的な哲学的問題に広がることはこれまでなかったように思われる。

BN の哲学的問題の中で、(少なくとも確率統計に少し関心のある私のような者にとって) 一際興味深いのは、モデル選択のスコアリング関数に関わる話である。BN は深層学習に対して数学的原理が明確とされつつ、最適モデルの選択に関しては様々な問題を抱えていて、スコアリング関数 (AIC, BIC, BDeu, fNML など) の選択もその一つである。これらは基本的に、いわゆるペナルティ項のみが異なるが、結果 (成績) がケースにより異なり、人工知能学者の支持もまちまちである。また研究メソッドはいずれも「経験的」であるように見える。この問題は一般的なモデル選択問題にもつながるが、BN に即して選択規準の選択をどこまで「原理的」に詰められるかは哲学的に関心が持たれるところである。

確率統計、人工知能、哲学が絡む興味深い問題はもちろん BN 単独の問題にとどまらない。近年の発展著しい深層学習だが、深層学習は「知覚」中心であり BN は「推論」中心であるという、ある種の棲み分け論が人工知能研究者の暗黙の了解であるように見受けられる。しかし一方で深層学習にベイジアンの考え方を組み入れようとする、哲学的好奇心をくすぐる試みもなされている (たとえば Gal & Ghahramani, 2015)。

哲学的問題だから哲学者が物を言える、というものでもないし、上記の問題はいずれもハードルが高い。では哲学者がこうした問題に関わりを持つとすれば、どんなアプローチの仕方があるのか。今回の発表では、それを探るための「行程表」のようなものを示したいと思う。

規則とパターン：後期ウィトゲンシュタインの洞察

松阪陽一（首都大学東京）

本発表では、ウィトゲンシュタインの有名な、「規則に従うこと」に関する議論を、新たな視点から理解することを目指します。ウィトゲンシュタインの規則遵守論に関しては、クリプキの解釈が有名です。クリプキの解釈は、ウィトゲンシュタイン研究者の間であまり評判のよいものではありませんが、他方、では「本当の」ウィトゲンシュタインの議論はどのようなものだったのかになると、クリプキを批判する人々の説明が、クリプキによる理解ほど哲学的に興味深いものであったかどうかは議論の分かれるところでしょう。クリプキの解釈の画期的であった点は、やはり、ウィトゲンシュタイン的なジャーゴンを最終的な説明項に用いずに、当時の分析哲学者に理解可能な概念を用いて、ウィトゲンシュタインの思想を説明しようとしたことにあると思われます。本発表で目指すのも基本的にはそのような試みで、その際、私は「パターン認識」の概念を中心に据えたいと考えています。これは近年認知科学や機械学習では盛んに研究されている概念ですが、分析哲学を含め伝統的な哲学はこの概念に適切な場所を与えておらず、ウィトゲンシュタインの議論がひどくわれわれに分かり難く感じられるのも、ひとつにはそのためではないかと私は推測しています。本発表が、この概念の哲学的重要性を少しでも明らかにできればと考えています。