

井上 克巳

INOUE, Katsumi

国立情報学研究所 情報学プリンシプル研究系 教授

[研究紹介]

■信頼性が十分とは言えない現在の AI

近年、人工知能（AI）に対する注目度が急激に高まっています。話題になっている ChatGPT など、プロンプトに応答してテキストや画像を作り出してくれる生成 AI は、誤情報やバイアスに左右されてとんでもない答えを返してることがよくあります。今のレベルでは、とうてい「信頼できる AI」とは言えません。

これまでの AI 研究の大きな流れとしては、以前は記号で表現された知識を扱う記号的 AI の研究が盛んでしたが、現在の主流はニューラルネットワークを中心とする機械学習による AI に移っています。その過程で、記号世界を扱う「知識表現・推論」と数値・ベクトル空間を扱う「機械学習」に分離され、それぞれの分野で独立して研究が行われる状況になっています。

計算機科学からのアプローチにより、1980 年代から一貫して AI の研究に携わってきた私としては、人が見て明らかに不適切であるような結果を AI が出力するような現状を一刻も早く改善したいと思っています。そのためには、AI の信頼性を向上させないといけません。不適切だったり首尾一貫しない回答をしないこと以外にも、想定外の変化やノイズに対する耐性（ロバスト性）などの要素も信頼性を得るためには考えていく必要があります。

■知識表現・推論と機械学習を統合する意義

私は、現在分離された状態になっている知識表現・推論と機械学習を融合することで、AI の信頼性を向上させることができると考えました。現在の研究を一言で表現すると「認識・学習・推論という一連の知能を共通の土台で扱うための基盤を提供」することになります。ここで「共通の土台」として考えているのが、代数的構造（行列・テンソル）を用いて論理式・制約を表現する代数的手法になります。

本来は記号処理（アルゴリズム）を使って解くべき推論問題を数値計算に置き換えることで、ロバストかつスケラブルな技術に進化させることが期待されます。私たちの代数的手法は、記号とサブ記号をともに扱う「ニューロシンボリック AI」と呼ばれる技術分野に属してはいますが、論理プログラムをテンソル表現する点で

オリジナルのアプローチに基づいています。

こうした私たちの研究を国際会議で発表する機会も増えており、2024 年度には 3 つの国際会議で招待講演の依頼を受けています。開発した研究成果は、論文発表以外にデータやプログラムもオープンソース化して公開していますが、これまで解けなかった問題が徐々に解けるようになり注目度も上がってきていることから、研究成果に確かな手ごたえを感じています。

■国際的な連携で「信頼できる AI」を目指す

現在、特に開発を進めたい技術は「制約を取り入れた AI」です。ここでいう制約とは記号知識として表現することを指しますが、制約知識を機械学習や生成 AI が取り入れることで、信頼性をどれだけ向上させることができるかという研究になります。

この研究の成果の一例として、2023 年に開催された自動運転における制約付き物体認識の性能を競う「ROAD-R challenge for NeurIPS」において、私たちのチームは、設定された 2 部門で優勝と 3 位という好成績を挙げることができました。この AI は、機械学習技術を駆使した物体認識技術に論理制約を要求として付加したもので、私の研究テーマである知識表現・推論と機械学習を融合させた応用例となっています。同様の考え方を生成 AI にも適用できれば、生成 AI の信頼性向上にもつながるはずですよ。

AI 研究において、日本は出遅れていると指摘する声もあります。私はこれまで、自然な流れで海外の研究機関や研究者と共同研究を行ってきました。これからも、オリジナルの研究を国際共同研究で推進して、信頼できる AI を目指していきます。それによって、日本の AI 研究が世界水準に追いつくことができると信じています。

AI はまだまだ未熟な点も多いかもしれませんが、各種システムの機能として実現される例が増えてきており、私たちは間接的に大きな恩恵を受けています。今後も AI が進化を続けていくことは間違いありませんので、私のように AI 基礎研究に取り組んでいる研究者を、ブームに左右されることなく見守っていただければ幸いです。☐

Researchmap <https://researchmap.jp/vivre?lang=ja>