

調査研究助成課題の成果概要(その1)

人工知能技術による学術研究促進に資する政策設計に関する国際比較研究

東京大学 未来ビジョン研究センター 教授
柴山 創太郎

1. 調査研究の目的

人工知能技術は、その中核を成す「機械学習」の性能の大幅な向上により、現代社会の様々な場面に応用されていますが、特に大学・公的研究機関において行われる学術研究においても機械学習は積極的に活用されており、科学技術進歩への貢献が期待されています。とりわけ、激化する国際競争の中で限られたリソースの下、科学者と機械の「協業」を促し、知識創造のプロセスを加速することは重要な政策課題と言えます。一方、学術研究に人工知能が「参画」することにより、科学者が担う役割や、科学者に求められるスキル・能力が変化し、人材育成や組織設計の在り方に修正が求められると考えられます。そこで、ドメイン科学(伝統的な学術分野)と計算機科学との学際的融合を促し、人工知能技術の活用を加速するための組織設計、及び、それを支える政策を検討することを目的としました。

2. 調査研究の方法

この目的を踏まえ、本プロジェクトでは主に2つの手法を採用しました。第1に、大規模データから一般的傾向を把握することを目的として文献書誌情報分析を採用しました。生物学、医学、農学、材料科学、化学、物理学の各ドメインにおいて、機械学習を活用した研究に取り組む2,500の研究チームを対象として、これらのチームから出版された論文を分析しました。これを「機械学習群」とし、さらに「対照群」として機械学習を活用していない22,500の研究チームから出版された論文を分析対象に加えました。第2に、大規模データ分析の結果を深掘りする目的で、機械学習群論文の責任著者を対象とした質問票調査を実施し、250名から回答を得ました。

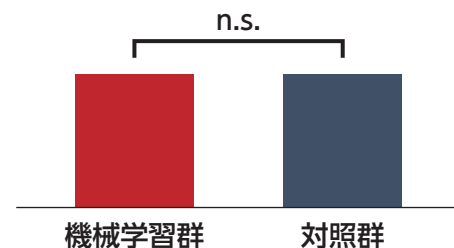
3. 調査研究の結果・考察

(1) 機械学習活用の効果

まず、科学者が機械学習を活用することにより期待する効用を質問票調査により確認したところ、「作業の効率化」(59%)、予測精度の向上(68%)、さらに、「人間の認知を超えた発見」(46%)などのメリットが

指摘されました。とりわけ「人間の認知を超えた発見」はより本質的な効用であると考えられたため、この点を文献書誌情報により確認しました。仮に「人間の認知を超えた発見」が重要であるとすれば、人が思い付くことの無いような斬新な発見が行われると予想されます。そこで、論文の新規性(Novelty)の指標について2群を比較してみましたが、有意な差は観察されませんでした(図1上)。一方、より一般的なメリットとして、研究成果のインパクトを論文の被引用件数に基づき比較したところ、機械学習群の方が有意にインパクトが高いことが示唆されました(図1下)。

(A) 情報の新規性指標(Novelty)



(B) 被引用件数(Impact)

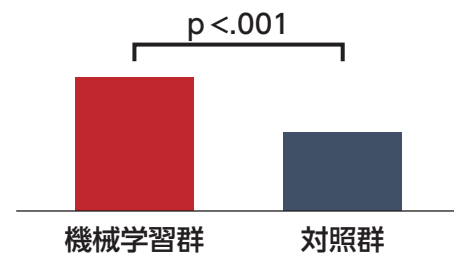


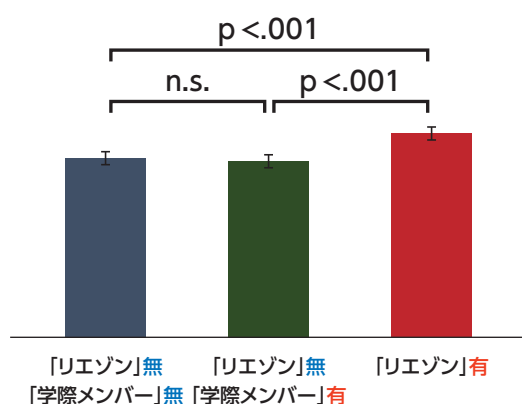
図1 研究成果の質的指標に対する機械学習の寄与
*出版年、出版誌、チームサイズを制御したうえで2群を比較。新規性指標はMatsumoto et al (2021)に基づく。

(2) 計算機科学とドメイン科学の橋渡し

次に、機械学習を活用する上での障壁について質問票調査の結果を分析したところ、「人材確保の難しさ」(42%)、「文化の相違」(26%)や、特に「計算機科学とドメイン科学の専門性に関する相互理解の不足」(68%)がボトルネックとして捉えられていることが

分かりました。そこで「相互理解の不足」を乗り越えて、計算機科学とドメイン科学を「橋渡し」するための組織的な対策として、①ドメイン科学と計算機科学のそれぞれの専門家を含む学際的メンバー構成、及び②ドメイン科学と計算機科学の双方に精通する研究者(所謂「リエゾン人材」)の採用について検討しました。分析の結果、リエゾンの存在はインパクト及び新規性の両方に寄与する一方、学際的メンバー構成は殆ど影響しないことが示唆されました(図2)。つまり、計算機科学とドメイン科学の両方に精通する人材(リエゾン)が、分野間の理解を媒介し、機械学習のメリットを引き出している可能性が考えられました。

(A) 情報の新規性指標 (Novelty)



(B) 被引用件数 (Impact)

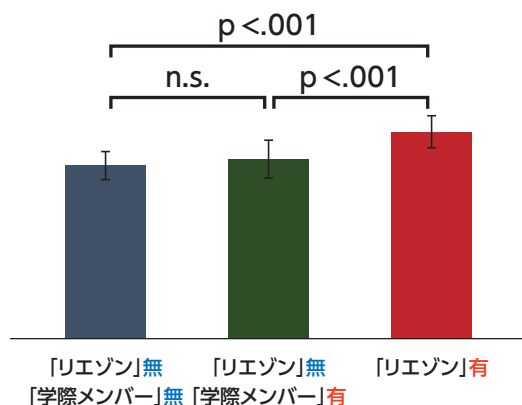


図2 計算機科学とドメイン科学を橋渡しする研究チーム
*出版年、出版誌、チームサイズを制御したうえで3群を比較。

(3) 国際比較

以上の論点を国間で比較してみます。まず機械学習関連論文のシェアを見ると、日本は4.3%と先進諸国中でも低いシェアに留まり(文献書誌情報分析サンプル中)、各国の全論文数で標準化した値で見ても、日本における機械学習の応用は先進諸国に比べて出遅れていることが分かりました(図3)。論文の質的指標で見ても、日本の機械学習論文は新規性が有意

に低いことが示唆されました(インパクトは有意差無し)。さらに日本の研究チームにはリエゾンの存在率が有意に低いことも示唆されました。これらの結果から、日本における人材面の制約が、機械学習応用の停滞の原因となっている可能性が考えられました。

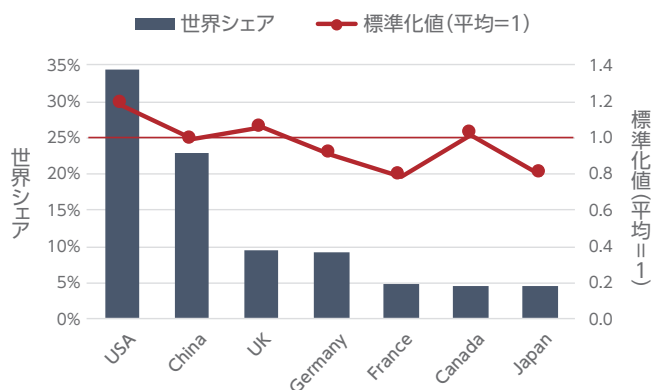


図3 機械学習活用論文のシェア

(4) 橋渡し人材の育成

最後に、リエゾンの育成が鍵を握るとの仮説から、人材育成の現状について質問表調査のデータを分析しました。既存のリエゾン人材の教育背景を見ると、大多数はドメイン科学の出身者が後に計算機科学に転向する形で育成されていることや、学際的分野(バイオ・インフォマティクス等)において早期から育成された人材は比較的少数であることが分かりました。また、ドメイン科学における計算機科学スキルの供給源を確認したところ、新規雇用や共同研究といった外部の供給源に比べて、内部で自前育成するとの回答が多数を占めました。これらの結果は、リエゾン人材の育成が、研究チームや個人のボトムアップの努力に依存している傾向を示唆するものであり、より体系的な人材育成の必要性を示唆するものです。このためには、計算機科学とドメイン科学との融合領域における研究・教育を促進する必要があると、とりわけ、かかる学際的課題に従事することに対して研究者個人に十分なインセンティブを与えることや、学際的なスキルを習得するためのキャリア支援が重要になると考えられました。

4. まとめ

以上の結果から、機械学習は学術研究の進歩に重要な貢献を果たし得ること、特にリエゾン人材は分野間の相互理解を助け、新規性の高い科学的発見を促進することが確認されました。日本ではリエゾン人材の不足が機械学習の活用の障害になっている可能性があり、そのような人材の体系的育成の重要性が示唆されました。分野間の理解不足や人材確保の難しさは世界共通の課題であり、学際的にオープンな環境を整備することで、機械学習の活用を通じた学術研究の一層の発展が期待されます。