

[Invited]Theoretical challenges to understand the principles of deep learning and artificial intelligence

*Masaaki Imaizumi¹

1. Graduate School of Arts and Science, the University of Tokyo

深層学習を用いたデータ解析の発展により、人工知能技術が発展し社会に変革をもたらしている。ただし、これらの技術には多くの課題も残されており、膨大化する計算コストや悪意ある攻撃に対する頑健性など、社会により浸透する中で改善すべき点も多い。

これらの課題を解決するための障害のひとつとして、深層学習・人工知能の原理の不理解がある。深層学習や人工知能は素晴らしい性能を発揮する技術だが、これらがなぜこのような性能を発揮できるかといった原理には未だ不明な点が多い。これらの不理解が、低コストな計算アルゴリズムや効率的な制御技術の開発を阻む一員となっている。よって、深層学習・人工知能を理解し効率的に改善するために、それらの仕組みを理解し記述するための理論基盤が必要とされている。

本講演では、(I)なぜ深層学習は高い精度を発揮できるのか、(II)人工知能をどのように数理的に表現・解析すれば良いか、という問いに関する理論的な研究を紹介する。

(I)の深層学習に関する研究では、特に (i)深層構造を持つモデルの優位性および劣位性とそれを改善する学習手法、(ii)深層モデルの不確実性を損失曲面の形状を用いて評価する理論、および(iii)大自由度モデルの漸近的な解析（過剰パラメータの理論）を深層構造に拡張する試みを紹介する。(II)の人工知能に関する研究では、特にChatGPTなどに代表される基盤モデル・大規模言語モデルに注目し、文脈内学習という汎用的な学習フレームワークの誤差解析を通じてその仕組みを理解するところを紹介する。