

人工知能の理解と制御に向けた深層学習ダイナミクスの解析：物理学的アプローチ

Analysis of Deep Learning Dynamics for Understanding and Controlling AI: An approach from Physics

*今泉 允聡^{1,2,3}

*Imaizumi Masaaki^{1,2,3}

1. 東京大学、2. 理化学研究所AIPセンター、3. 京都大学

1. The University of Tokyo, 2. RIKEN AIP, 3. Kyoto University

人工知能の発展に伴い、それを構成する深層学習の内容を理解する理論も発展している。特に、深層学習の予測性能を説明する上で学習ダイナミクスの役割が注目されているが、ニューラルネットワークの多層構造で現れる学習ダイナミクスの解明は未だ発展途上である。本講演では、ニューラルネットワークの動力学を解析する幾つかの研究成果を紹介する。第一に、ニューラルネットワークの層ごとに異なる時間構造のズレが、学習の効率性に影響を与えるという理論的結果を報告する。第二に、層の多いようなニューラルネットワークにおいても、その学習ダイナミクスおよび予測誤差を記述・推定する技術について報告する。さらに時間があれば、ニューラルネットワーク、特にトランスフォーマーの内部で自律的に実現する変数分布のパターンについて紹介する。これらの解析は、統計力学や複雑系物理学の成果を機械学習に導入することで得られる。