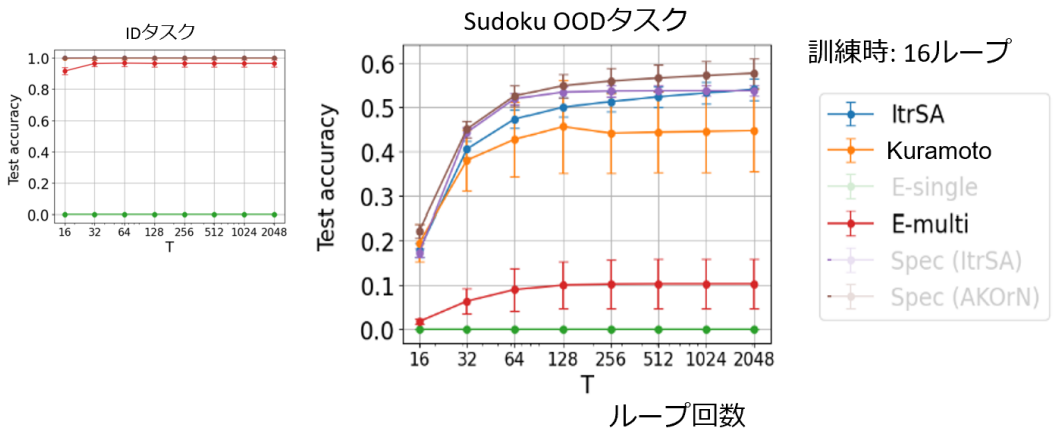
ABCI 3.0開発加速利用 （2025年度）成果概要（公開用）

課題名：
深層学習の数理と応用のための数値実験

実施時期：2025年4月-2026年3月
所属機関名：産業技術総合研究所
代表者氏名：唐木田亮

成果概要：
自己注意機構の統計物理学的モデリングを検討し、理論の妥当性を数値実験検証した。具体的には、(1) エネルギー関数を持つ自己注意機構モデルの構成方法を提案し数値的に検証、(2) コンテキスト長が大きい極限でのランク崩壊を防ぐ逆温度設定の実験による検証、を行った。

成果のポイント：
成果(1)に関しては、既存研究に比べより現実に近い設定で、エネルギー関数(リアプノフ関数)を構成した。具体的には、対称性の仮定を緩め、さらにマルチヘッドへの拡張を行った。このエネルギー関数がある再帰自己注意モデルは安定した状態更新を実現すると期待されるが、実験により必ずしもこの安定性が推論に優位ではないことが明らかになった (右図)。予測時はID (分布内)データではエネルギー関数の保証を促す正則化ありのモデル (E-multi)が高い性能を達成できるているが、OOD(分布外)データではエネルギー保証なしの系に比べ劣っていることがわかる。これは固定点に収束するようなダイナミクスではなく、振動のようなより動的な状態変化が望ましいことを示唆している[詳細 論文1]。
成果(2)に関しては、コンテキスト長が大きい極限で、ランク崩壊を回避する逆温度のスケールが存在することを数値実験で確かめることができ、理論解析の指針として機能した [詳細 論文2]。



成果についてより詳細な情報を提供しているWebページ、発表論文などの情報：
論文1: "Recurrent Self-Attention Dynamics: An Energy-Agnostic Perspective from Jacobians", Akiyoshi Tomihari, Ryo Karakida, NeurIPS 2025
論文2: "Gaussian Equivalence for Self-Attention: Asymptotic Spectral Analysis of Attention Matrix", Tomohiro Hayase, Benoît Collins, Ryo Karakida, AISTATS 2026