

グラフや順序をもつ集合など、 さまざまな構造を取り扱うための学習とは？ さまざまな構造と学習

杉山 磨人 (国立情報学研究所 情報学プリンシプル研究系)

どんな研究？

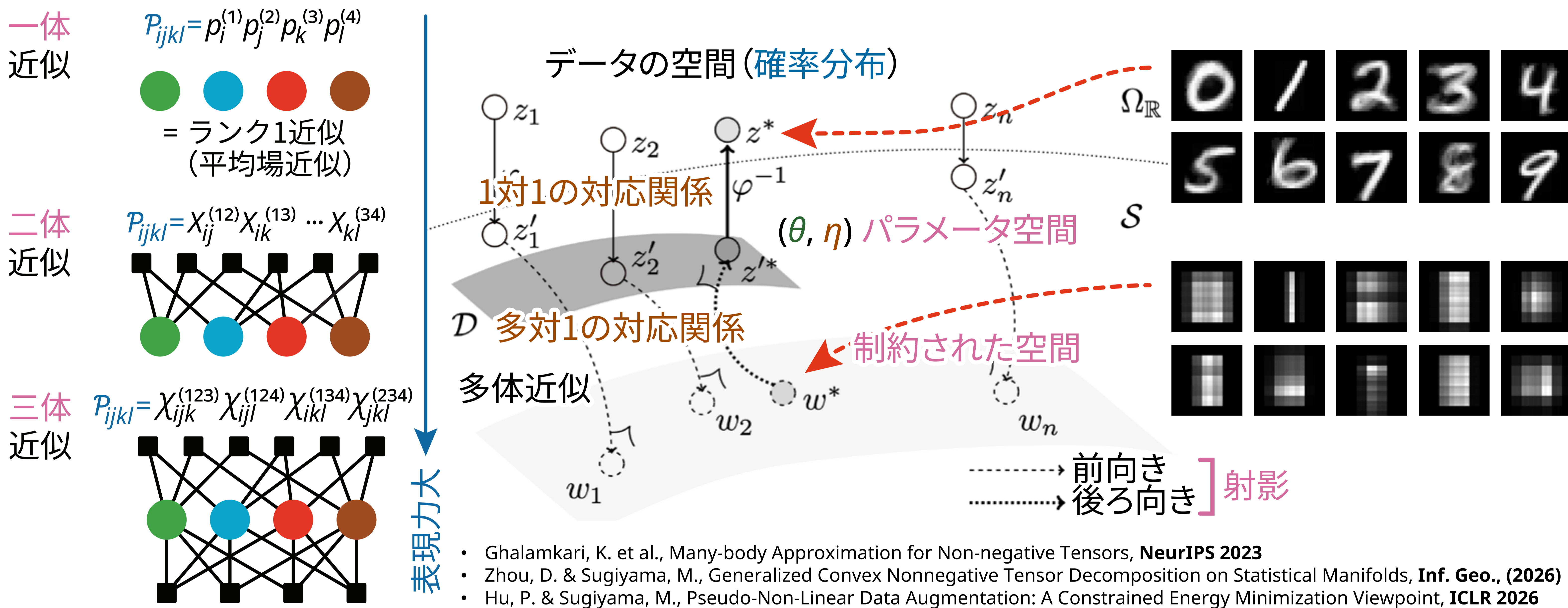
- 機械学習では、多様な構造を持つデータや対象が現れる
- これらの構造は、明示的なものから潜在的なものまで、さまざまな形で存在している
- そのような構造を効果的に活用する学習手法の開発を目指す

何がわかる？

- 機械学習手法の表現能力や限界を理論的に理解する
- 多様な構造を効果的に扱うための学習技術の基盤を構築する
- 現実世界に存在する多様な構造を適切に数理モデル化する方法論を確立する

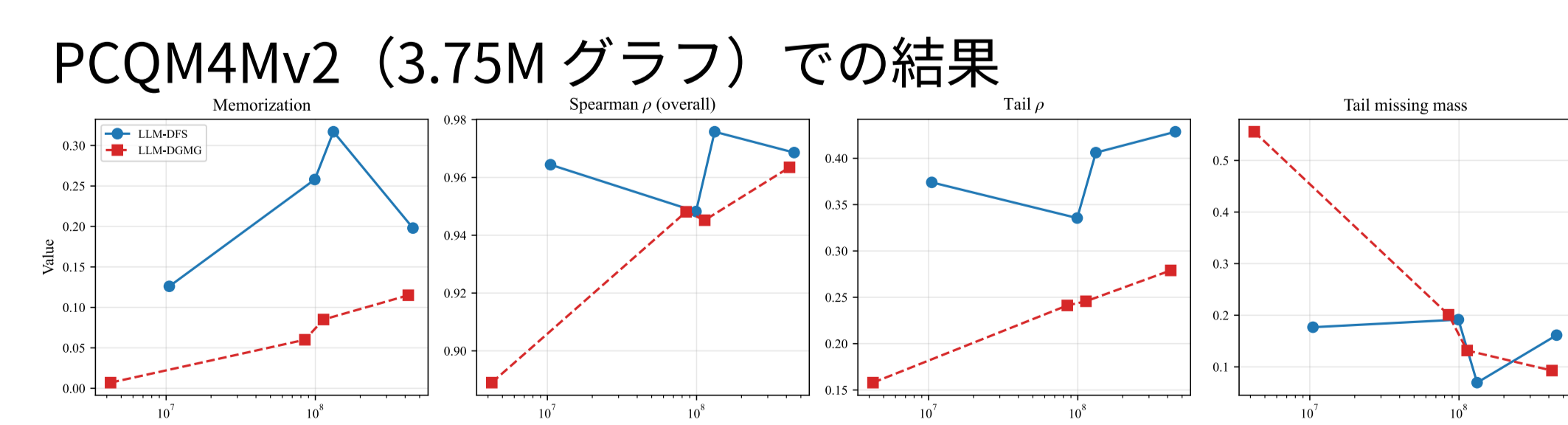
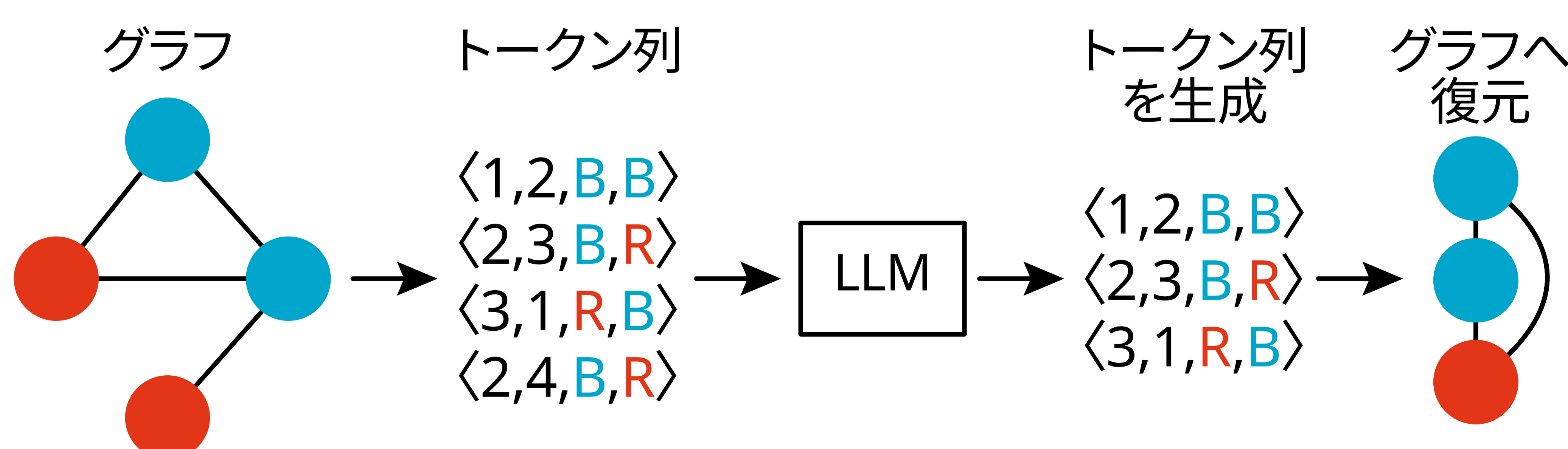
テンソル多体近似によるデータ拡張

- テンソル (多次元配列) を分解する手法である多体近似を開発し、その手法を用いたデータ拡張手法を提案



言語モデルによるグラフ機械学習

- グラフをトークン列へと系列化し、LLMで学習し、そこから新しいグラフを生成する **グラフ言語モデル**
- グラフをトークン列へとエンコードする手法として、Canonical DFS codeとDGMG action sequenceを採用
- **知見1**：グラフデータのサイズが小さいとき (~10万グラフ程度)、LLMはグラフを完全に記憶することができる
- **知見2**：グラフデータのサイズが大きくなると (約400万グラフ)、LLMはグラフの丸暗記はできなくなるが、出現頻度が高いグラフ部分構造を生成することができる。つまり、LLMはグラフ構造を学習できている



- Yamada, M., Sugiyama, M., When Graph Language Models Go Beyond Memorization, **arXiv:2605.06239**
- Larsen, L., et al., Same Graph, Different Likelihoods: Calibration of Autoregressive Graph Generators via Permutation-Equivalent Encodings, **AISTATS 2026 WS**