

2023 年度年次報告書

数学・数理科学と情報科学の連携・融合による情報活用基盤の創出と社会課題解決に向けた展開

2021 年度採択研究代表者

田中 久美子

早稲田大学 理工学術院
教授

自然言語の非線形性の計算論モデル

主たる共同研究者:

峯島 宏次 (慶應義塾大学 文学部 准教授)

宮尾 祐介 (東京大学 大学院情報理工学系研究科 教授)

研究成果の概要

2022 年度後半に公開された大規模言語モデル(LLM)に、本プロジェクトは影響を受け、計画を大幅に見直すことが必須となった。一部の研究を終了とし、「統計」と「論理」の観点から体制を整理し、残期間 3 年に向けた軌道修正を行なった。現行の LLM の限界を受け、統計・論理班が協働し、自然言語の特質を数理的に探求することを通し、信頼性の高い知的な AI を目指すものとして整理された。見直しを 3 年次の後半に行ったことは、本研究プロジェクトを健全に進める上で、一つの大きな前進であると考え。前半は、基礎的な個別研究に取り組む計画で、それは順調に進んでいる。

複雑系班は、マクロ 1 (系列の非線形性と数理モデル)、メゾ 1 (文構造の数理モデル)、マイクロ 1 (単語・文の埋め込み表現)に取り組んでいる。非線形性の観点から、木構造・多峰性を軸として、ユークリッド空間を超える空間で、言語対象を考察することを試みた。たとえば、マクロ 1 では、力学系解析の一つの方法である相関次元の計測を行った。また、メゾ 1 では、文構造の考察を行うために、ストラ数という、系の分岐度合いを計数する複雑系の概念を転用した。

計算言語班は、メゾ 1 (文構造の非線形性の数理モデルと構文解析への応用)とマイクロ 2 (単語構造から文構造までの統合的数理モデル)について、前年度に引き続き研究を続けた。

数理論理班は、マクロ 2 (文列の構造の特性)、メゾ 2 (論理表現に基づく文の推論)、マイクロ 2 (単語・文の論理式への変換可能性)に関する研究を進めた。マクロ 2 は、範疇文法のモデルであるランベック計算とその拡張系の計算の複雑さを解明する研究を発展させた。メゾ 2 については、LLM の論理推論能力を評価する研究を進めた。マイクロ 2 については、自然言語の木構造から、高階論理式への変換を行うモデルの整備・拡張を進めた。

【代表的な原著論文情報】

- 1) Xin Du and Kumiko Tanaka-Ishii. Correlation dimension of natural language in a statistical manifold. *Physical Review Research*, 6, L022028, 2024.
- 2) Kumiko Tanaka-Ishii and Akira Tanaka. Strahler number of natural language sentences in comparison with random trees. *Journal of Statistical Mechanics*, 2023(1234034), 2023. Selected for JSTAT Scientific Directors for the Highlights collection 2024.
- 3) Taiga Ishii, Yusuke Miyao. Tree-shape Uncertainty for Analyzing the Inherent Branching Bias of Unsupervised Parsing Models. *Proceedings of the 27th Conference on Computational Natural Language Learning (CoNLL)*, pages532–547, 2023 年 12 月
- 4) 谷中瞳, 峯島宏次 「AI は言語の基盤を獲得するか: 推論の体系性の観点から」『認知科学』31 巻 1 号, 27-45, 2024 年 3 月.
- 5) Kentaro Ozeki, Risako Ando, Takanobu Morishita, Hirohiko Abe, Koji Mineshima, and Mitsuhiro Okada, Exploring Reasoning Biases in Large Language Models Through Syllogism: Insights from the NeuBAROCO Dataset. *Findings of The 62nd Annual Meeting of the Association for Computational Linguistics (ACL2024)*, 2024 年 8 月.