

2024 年度年次報告書

次世代 AI を築く数理・情報科学の革新

2024 年度採択研究代表者

芳賀 あかり

奈良先端科学技術大学院大学 先端科学技術研究科
大学院生

乳幼児の発話誤りの模倣に基づくマルチモーダル言語獲得モデル

研究成果の概要

本研究は、子供が少ないデータで言語を習得することに着目し、大量の学習データを必要とするマルチモーダル言語モデルの学習において、子供が示す典型的な「誤り」をあえて模倣することで、モデルの学習効率向上を目指すものである。本研究では、特に視覚的な情報と言葉を結びつける際の子供の誤りの傾向を模倣することを目的とする。第一年度は留学のため実装までは至らなかったが、先行研究の整理と実験の設計を進めるとともに、関連の研究を行うフローニンゲン大学の Arianna Bisazza 研究室にて、本研究に必要な周辺知識を学んだ。さらに、同研究室との共同研究として、子供が言語習得過程で受け取るインプットの特徴が言語モデルの学習に与える影響について調査を行った。具体的には、子供向けの発話によく含まれるとされる「バリエーションセット」と呼ばれる発話集合が言語モデルの学習に与える影響を分析した。その結果、バリエーションセットが言語モデルの文法構造の習得や自然言語理解に有用であることを示した [成果 1]。これらの成果により、言語処理学会第 31 回年次大会において若手奨励賞を受賞した。

【代表的な原著論文情報】

- 1) 芳賀 あかり, 深津 聡世, 大羽 未悠, Arianna Bisazza, 大関 洋平. 言語モデルの事前学習におけるバリエーションセットの効果. 言語処理学会第 31 回年次大会 (2025).