

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

2022 年度
年次報告書

田中 久美子

東京大学 先端科学技術研究センター
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

マクロ1では、力学系解析の一つである相関次元解析を、自然言語の高次元埋め込み表現の系列に対して適用した。ランダム列よりも複雑な自然言語の特質が明らかになると共に、現行の埋め込み表現の限界が相関次元により計測できることが明らかとなった。

ミクロ1については、既存の埋め込み表現を再考した。既存の埋め込み表現の表現力を高めるために、多峰性を表現可能な関数による埋め込み表現を実現し評価した。

メゾ1については、構文木の非線形性を計量する手法を提案し、既存モデルが右分岐の木構造を生成するバイアスを持つことを明らかにした。また、言語モデルの統計量を用いた教師なし構文解析モデル、構文構造に基づく言語横断意味解析モデルの研究に着手した。

ミクロ2については、これまで整備を進めてきた多言語組合せ範疇文法ツリーバンクの精緻化を進め、多言語構文解析の研究に利用できる環境を実現した。GitHubにおいて公開した。

メゾ2については、組合せ範疇文法から論理式への変換を行う意味解析器の構築・改良を進めると同時に、多様な言語現象を考慮した汎用的な意味割り当ての手法を実現するための研究を立ち上げた。範疇文法の数理論理的モデルとして重要なランベック計算とその拡張系に着目し、その計算量を解明する一連の成果を得た。また、比較表現・数量表現を含む多様な文を論理式に構成的に変換する意味解析器を構築し、自動定理証明に基づく論理推論により評価を行った。

マクロ2については、既存の文列の談話関係・含意関係コーパスの利用、新たなコーパス構築の両面から次年度以降の基盤となる研究に着手し、文列の処理に向けた数理論理的なモデルの検討を進めた。現状の深層学習モデルが自然言語の含意関係にどの程度汎化するのかを精査し、現状のモデルでは自然言語の構成性を反映した含意関係に弱点があることを明らかにした。

【代表的な原著論文情報】

- 1) Junjie Chen, Xiangheng He, Yusuke Miyao, “Modeling Syntactic-Semantic Dependency Correlations in Semantic Role Labeling Using Mixture Models”, Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pp.7959-7969, 2022.
- 2) Tu-Anh Tran, Yusuke Miyao, “Development of a Multilingual CCG Treebank via Universal Dependencies Conversion”, Proceedings of the Thirteenth Language Resources and Evaluation Conference (LREC 2022), pp.5220-5233, 2022.
- 3) Izumi Haruta, Koji Mineshima, Daisuke Bekki, “Implementing Natural Language Inference for Comparatives”, Journal of Language Modelling, Vol.10, No.1, pp.139-191, 2022.
- 4) Hitomi Yanaka, Koji Mineshima, “Compositional Evaluation on Japanese Textual Entailment and Similarity”, Transactions of the Association for Computational Linguistics (TACL) Vol.10, pp.1266-1284, 2022.
- 5) Hiromi Tanaka, “Tower-Complete Problems in Contraction-Free Substructural Logics”, Proceedings of 31st EACSL Annual Conference on Computer Science Logic (CSL 2023), pp.34:1-34:19, 2023.
- 6) Xin Du and Kumiko Tanaka-Ishii. Semantic Field of Words Represented as Nonlinear Functions.

36th Conference on Advances in Neural Information Processing Systems (NeurIPS 2022), 2022.