

2023 年度年次報告書

文理融合による人と社会の変革基盤技術の共創

2022 年度採択研究代表者

坂地 泰紀

北海道大学 大学院情報科学研究院

准教授

因果情報を用いた経済ナラティブシミュレーション

研究成果の概要

本テーマは、因果関係インスタンスを用いて因果チェーンを構築し、イベントが発生した時の因果チェーンの変化を予測することで、ナラティブの変化が経済へどのような影響をもたらすかシミュレーションすることで、ナラティブ経済学を実証することを目的としている。

現在、日本語と英語に関する因果関係認識技術については、概ね確立できていると言える状況にある。グラフニューラルネットワークと構文解析技術を組み合わせることで、日本語データセットにおいても、英語データセットにおいても高い性能で因果関係を認識できる手法を開発した。こちらに関しては、査読付き国際ジャーナルに投稿していたところ、無事に採録された。

本テーマでは、因果チェーンを用いて様々なテーマでの経済ナラティブを認識しようと研究を進めており、まず、テーマを気候変動に絞り、かつ、因果チェーンも多層ではなく2層のものに絞ることで、気候変動ナラティブを把握する研究を進めている。こちらの研究は、日本銀行との共同研究になっている。こちらの研究を進めたものが査読付き国際会議に採択され、イタリアのソレントにて研究発表を行った。

2022年からChatGPTを代表する生成AIが世間を圧巻するようになった。新しい生成AIによって、当初立てていた研究計画を見直し、大規模言語モデルの活用可能性の検討を新たに始めた。2023年当初は、日本語特化型の大規模言語モデルが存在しなかったことから、公開されている大規模言語モデルを日本語へチューニングする試みを行った。結果的には、当初から日本語のテキストを学習させた大規模言語モデルが必要であるという結論になったが、この取り組みをする中で作成したデータセットを公開するのに加え、途中経過をいくつかの国内学会で発表したところ、こちらが高く評価され、電子情報通信学会研究会の優秀研究賞を頂くことができた。

【代表的な原著論文情報】

- 1) Hiroki Sakaji, Kiyoshi Izumi, "Financial Causality Extraction based on Universal Dependencies and Clue Expressions," *New Generation Computing*, <https://doi.org/10.1007/s00354-023-00233-2>, 2023.
- 2) Masahiro Suzuki, Hiroki Sakaji, Masanori Hirano, Kiyoshi Izumi, "Constructing and analyzing domain-specific language model for financial text mining," *Information Processing & Management*, Vol.60, No.2, <https://doi.org/10.1016/j.ipm.2022.103194>, 2023.
- 3) Hiroki Sakaji, Noriyasu Kaneda, "Indexing and Visualization of Climate Change Narratives Using BERT and Causal Extraction", 2023 IEEE International Conference on Big Data (IEEE BigData 2023), pp.5674-5683, Sorrento, Italy, December, 2023.
- 4) 鈴木雅弘, 平野正徳, 坂地泰紀, 日本語インストラクションデータを用いた対話可能な日本語大規模言語モデルのLoRAチューニング, 第20回テキストアナリティクス・シンポジウム, 2023.
- 5) 平野正徳, 鈴木雅弘, 坂地泰紀, llm-japanese-dataset v0: 大規模言語モデルのための日本語チャットデータセット構築, 情報処理学会 第257回自然言語処理研究発表会, 2023.