

日独仏 AI 研究

2020 年度採択研究代表者

2022 年度

年次報告書

松本 裕治

理化学研究所 革新知能統合研究センター

チームリーダー

医薬品安全性監視のための言語を超えた知識強化情報抽出

主たる共同研究者:

相澤 彰子 (情報・システム研究機構 国立情報学研究所 教授)

荒牧 英治 (奈良先端科学技術大学院大学 先端科学技術研究科 教授)

研究成果の概要

医薬品安全監視に関する SNS サンプルに対するアノテーション作業を実施して問題点の洗い出しを行い、独仏グループとのオンラインミーティングを通じてアノテーション仕様の詳細化を行った。SNS サンプルに対してアノテーション作業を進めながら、仕様の修正作業を実施した。

十分な訓練データの入手が困難な状況での固有表現認識、および、関係解析技術に関する手法を構築した。また、固有表現および関係情報抽出を同時に生成する生成モデルを提案し、その有効性を確認した。また、Covid-19 に関する学術論文データを対象として、新規ドメインにおける固有表現抽出手法の評価と誤り分析を行い、国際会議で発表した。

医学分野の日本語論文の抄録から事前学習済モデル (RoBERTa) を構築して公開した。また、ベンチマークデータセットを使って構築した言語モデルの有用性を評価し、論文を対象とする言語タスクにおいて従来のモデルより優れた性能が得られることを示した。さらに、論文中に出現する文の機能に注目した埋め込み表現の獲得手法を提案して、国際ワークショップで発表した。

Tweet 約 2000 件をサンプリングし、症状や医薬品の表現と、有害事象の有無をアノテーションしたデータを独仏チームに先行して構築した。また、第 1 版として完成したアノテーション仕様に基づいて、新規ツイート 560 件および質問投稿サイト (Yahoo!知恵袋) の副作用に関する質問・回答 60 件からなるコーパスのアノテーションを完了した。

医薬品安全に関する国際シェアードタスク NTCIT-17 Real-MedNLP で用いた、副作用情報が付与された日英パラレルコーパスを中国語にも翻訳し、国際的可用性を高めた。また、副作用に関する Tweet を元に言語生成モデルで出力した 10,000 件の擬似 Tweet (日本語) に副作用情報を付与したデータを作成した。

【代表的な原著論文情報】

- 1) Van-Hien Tran, Hiroki Ouchi, Taro Watanabe, Yuji Matsumoto, “Improving Discriminative Learning for Zero-Shot Relation Extraction,” Proceedings of the 1st Workshop on Semiparametric Methods in NLP: Decoupling Logic from Knowledge, pp.1-6, Dublin, Ireland and Online, May 27, 2022.
- 2) An Dao, Akiko Aizawa and Yuji Matsumoto: “Named Entity Recognition on COVID-19 Scientific Papers,” 2022 6th International Conference on Natural Language Processing and Information Retrieval (NLP4IR 2022), Bangkok, Thailand. December 16-18, 2022.
- 3) Kaito Sugimoto and Akiko Aizawa; “Incorporating the Rhetoric of Scientific Language into Sentence Embeddings using Phrase-guided Distant Supervision and Metric Learning,” Third Workshop on Scholarly Document Processing (SDP 2022), Workshop at COLING 2022. Gyeongju, Korea. October 17, 2022.
- 4) Lisa Raitel, Faith W. Mutinda, Gabriel H. B. Andrade, Hui- Syuan Yeh, Tomohiro Nishiyama, Mathieu Laï-King, Shuntaro Yada, Roland Roller, Cyril Grouin, Agata Savary, Aurélie Névéal, Thomas Lavergne, Eiji Aramaki, Sebastian Möller, Yuji Matsumoto, and Pierre Zweigenbaum: “KEEPHA at n2c2 2022: Track 1 Contextualized Medication Event Extraction,” 2022 n2c2 Shared Task and Workshop, 2022