

AI 活用で挑む学問の革新と創成
2020 年度採択研究者

2021 年度 年次報告書

中山 功太

筑波大学 理工情報生命学術院システム情報工学研究群
大学院生(博士後期課程)

31 言語における FG-NER・EL システム開発

§ 1. 研究成果の概要

本研究で開発する FG-NER・Entity Linking(EL)システムは主に、文章中の固有表現の言及を検出する言及検出、言及を各固有表現クラスに分類する言及分類、言及を Wikipedia などの知識ベースのエンティティに紐付けるエンティティリンキングからなる。これらシステムを学習するためにはデータ作成に膨大なアノテーションコストを必要とするため、本研究では Wikipedia に存在する内部リンクを用いて、低コストに学習を行う手法の開発を行う。達成すべきタスクとして、各エンティティは事前に固有表現クラスに分類されている必要がある。分類は新羅プロジェクトにより開催された共有タスクにおいて既になされているが、各エンティティに対して複数のタスク参加者より複数のラベルが付与されており、統合する必要がある。そのため、共同教育といった手法を提案した¹⁾。本手法では、参加システムの予測結果をもとに、単一のシステムで各参加システムを再現し、同時に、共有タスクで使用された学習用データで最終出力層を学習することで統合結果を予測可能なシステムを得る。本手法の利点は、参加者の予測結果とタスクで使用された学習データが得られれば動作可能である容易さと、統合システム自体を再現するため、入力データが増えた場合や、入力データに何らかの変更があった場合にも、元の参加システムを動作させることなく新しい予測結果を得ることができる点である。また、昨年度は言及検出とエンティティリンキングの手法開発を行なったが、本年度の実験の中で、知識ベースに含まれないエンティティである NIL の検出精度が不十分であることが分かった。NIL 検出精度を改善する必要があるが、実用可能な評価セットが存在しない問題がある。そのため、本年度は、Wikipedia ダンプを用いた NIL 検出評価セット作成方法を考案し、現在、本評価セットを用いた高精度な NIL 検出手法の開発を行なっている。本成果は来年度に国際学会に投稿予定である。

【代表的な原著論文情報】

1) Kouta Nakayama, Shuhei Kurita, Akio Kobayashi, Yukino Baba, and Satoshi Sekine. 2021. [Co-Teaching Student-Model through Submission Results of Shared Task](#). In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 4525-4535, Punta Cana, Dominican Republic. Association for Computational Linguistics.