

AI 活用で挑む学問の革新と創成
2020 年度採択研究者

2020 年度 年次報告書

中山 功太

筑波大学 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、文章中の固有表現の言及 (Mention)を検出する Mention Detection(MD)、Mention を各固有表現(NE)クラスに分類する Mention Classification、Mention を Wikipedia などの知識ベースの Entity に紐付ける EL に分けることができる。

本年度は、MD と EL における基礎的な技術開発にあたった。特に MD はその技術を人手作成の学習データを用いた深層学習モデルや、Mention の表層文字列の辞書を用いたマッチングに頼っている。しかし、人手アノテーションのコストは特に他言語適用する場合には無視できないものとなる。また、表層辞書のマッチングは同音異義語といった曖昧性を持つ Entity を解消できないといった問題がある。そのため、多くの研究が Wikipedia 記事に含まれる内部リンクを用いて問題解決にあたっているが、編集ガイドラインの都合上、記事内で既にリンクしている Entity に関するリンクや自己を参照するリンクは付与されていない。

以前の研究では表層マッチングによりリンク補完を行なっているが、依然として曖昧性の問題が残る。我々は編集ガイドラインに則って深層学習モデルを学習することで、リンクが不十分な記事からの MD・EL システムの構築を可能とする。具体的には、学習段階ではモデルは対象の Mention が知識ベース内に含まれているか、既に出現しているかといった情報を受け取り、モデルは知識ベースに存在する最初の出現の検出を行うが、評価時に全ての Mention が出現していない状況としてモデルを作動させることで、2 回目以降の Mention も検出できるといった仕組みである。現状評価データがないため、正答データのアノテーションと正確な評価、成果公開は来年度の課題とする。