

AI 活用で挑む学問の革新と創成
2022 年度採択研究代表者

2022 年度
年次報告書

河野 誠也

理化学研究所 情報統合本部
特別研究員

特許文書読解支援のための談話構造解析基盤の確立

研究成果の概要

本研究では、特許文書の文間の参照・被参照関係を、知財的視点からきめ細やかなレベルで同定するための談話構造解析モデルを実現し、当該発明における技術的範囲を理解・分析するための解析基盤を構築することを目的とする。本年度は、(1)特許文書における請求項の構造に着目し、請求項間の参照・被参照関係を解析するためのアノテーション済みデータセットと解析モデルの構築を進めた。具体的には、特許庁から過去数十年分の特許情報を取得し、アノテーション対象とする小規模な特許集合(1000件)を選定した。特許集合の選定においては、特許登録の有無や技術分野の多様性に特に留意した。次に、選定した特許集合における各請求項間の参照・被参照の有無を半自動でアノテーションした。さらに、それらの参照・被参照がどのような関係を持つかのラベルを知財的視点から複数定義し、アノテーションを進めた。また、このような文間の関係性を同定するようなベースラインモデルとして、深層学習に基づいた談話構造解析モデルを構築し、その評価を実施した。ただし、評価用データセットの構築は進行中であるため、特許ドメイン以外の公開コーパスを用いて評価した。また、次年度以降に実施する請求項構造解析モデルの有用性の検証のための準備として、(2)類似特許の検索タスク、特許請求項の書き換えタスクに着目し、これらのタスクのための評価用データセットを構築した。このようなデータセットの構築においては、特許庁による公開情報を用いて半自動で構築が可能であることに着目した。具体的には、特許検索タスクについては、特許審査官がある特許を審査する際にその特許を拒絶する理由として引用した特許情報が正解データとして利用できる。また、特許請求項の書き換えタスクについては、特許審査官による特許の拒絶理由通知の内容に応じて、請求項の内容を修正することができる。これは、拒絶された特許をどのように書き換えれば特許が成立するかを評価するための正解データとして利用できる。いずれの、タスクにおいても、特許請求項の構造の深い理解は必要不可欠であり、本研究で構築する解析モデルの有用性の検証に適していると考えている。

【代表的な原著論文情報】

1) Seiya Kawano, Koichiro Yoshino, David Traum, Satoshi Nakamura, "End-to-End Dialogue Structure Parsing on Multi-Floor Dialogue Based on Multi-Task Learning", *Frontiers in Robotics and AI*, <https://doi.org/10.3389/frobt.2023.949600>, 2023.