

**厚生労働科学研究費補助金**  
**(政策科学総合研究事業 (臨床研究等 ICT 基盤構築・人工知能実装研究事業))**  
**分担研究報告書**

課題名 : 新薬創出を加速する症例データベースの構築・拡充/創薬ターゲット推定アルゴリズムの開発に関する研究  
研究分担者名 : 高村 大也  
国立研究開発法人産業技術総合研究所・人工知能研究センター・研究チーム長

**研究要旨**

基礎生物学系文献からの分子、病態の知識抽出については、構築した言語リソースの評価を進め、またそれを用いて、エンティティ、関係及びイベント抽出モデルの開発及びツール化を行った。また、論文から情報を抽出するキュレーション技術の開発を行った。また、疾患ネットワークの構築について、その方法の開発及び、視覚化及びツール化を進めた。

**A. 研究目的**

本研究プロジェクトは、論文などの技術文献から、生命系のエンティティ（蛋白質、疾患など）やイベント、及びそれらの間の関係を、自動的に抽出する技術を開発することを目的とする。さらに、それを用いて実際に知識抽出を行い、抽出した知識を、エンティティをノード、イベントをエッジとし、関係を持つノード同士を結合した、巨大な高次ネットワーク構造（疾患ネットワーク）で表す。

より具体的には、次の3つの課題を解く。

- ① 基礎生物学系文献からの知識抽出
- ② 基礎生物学情報と抽出情報との融合
- ③ コーパスデータの拡充

**B. 研究方法****① 基礎生物学系文献からの知識抽出：**

前年度までに開発した、文献からのデータキュレーション技術の評価、およびまとめを行い、その内容を論文として発表した。また、コードや環境の整備を行い、ウェブアプリケーションとして公開した。

**② 知識の表現方法及び柔軟な推論技術の開発：**

前年度までに開発した文献解析技術の結果を統合し疾患ネットワークを構築する技術を開発した。特に、結果の統合の際の規則の策定と実装を行なった。構築した技術はウェブアプリケーションとして公開した。またその内容を論文としてまとめ投稿した。

**③ コーパスデータの拡充：**

IPF基礎分子系コーパス構築のために策定したアノテーションガイドラインとアノテーションスキームを利用した、肺がん基礎分子系の文献へのアノテーションを行なった。

**(倫理面への配慮)**

本研究は、人及び動物を研究対象としていない研究であるため、倫理面の問題はないと判断した。

**C. 研究結果**

データキュレーション技術について、評価を行い、ウェブアプリケーションとして実装し公開した。文献を入力することで、事前に指定された情報が自動的に抽出される。また、疾患ネットワークの自動構築技術については、文献解析結果の統合ルールを策定および実装し、ウェブアプリケーションとして実装した。文献を入力することで、そこから疾患ネットワークを自動的に構築することができる。結果は、2次元および3次元で視覚化できるようになっており、またネットワークの構造をデータとしてダウンロードすることも可能になっている。コーパスデータの拡充については、肺がん基礎分子系の文献の抽象化100件のアノテーションを行なった。

#### D. 考察

データキュレーション技術および疾患ネットワーク構築技術については、ウェブアプリケーションという形で実装できた。これらのアプリケーションを安定して運用していくためのメンテナンスを必要とする。

#### E. 結論

データキュレーション技術、疾患ネットワーク構築技術、コーパスデータの拡充について、予定していた開発目標を達成した。

#### F. 研究発表

##### 1. 論文発表

- 1) Nozomi Nagano, Narumi Tokunaga, Masami Ikeda, Hiroko Inoura, Duong A. Khoa, Makoto Miwa, Mohammad G. Sohrab, Topic Goran, Mari Nogami-Itoh, Hiroya Takamura, “A novel corpus of molecular to higher-order events that facilitates the understanding of the pathogenic mechanisms of idiopathic pulmonary fibrosis”. Scientific Reports, (accepted) 2023.

##### 2. 学会発表

- 1) Mohammad Golam Sohrab, Khoa N.A. Duong, Ikeda Masami, Goran Topic, Yayoi Natsume-Kitatani, Masakata Kuroda, Mari Nogami Itoh and Hiroya Takamura. “BiomedCurator: Data Curation for Biomedical Literature”. In Proceedings of ACL-IJCNLP, Systems Demonstrations, pages 63-71, 2022.
- 2) Mohammad Golam Sohrab, Anh-Khoa Duong Nguyen, Goran Topic, 伊藤眞里, 黒田正孝, 夏目やよい, Masami Ikeda, 高村 大也. “BiomedCurator: 医学・生物学文献からの構造化データ 抽出のためのデータキュレーションシステムの開発”. 言語処理学会年次大会, 2023.

#### G. 知的財産権の出願・登録状況（予定を含む。）

##### 1. 特許取得

なし

##### 2. 実用新案登録

なし

##### 3. その他

なし