

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知識抽出については、前年度までに開発し公開したデータキュレーションシステムについて、メンテナンスを行い、サービス提供を継続した。知識の表現方法及び柔軟な推論技術の開発については、疾患ネットワークを構築し視覚化するシステムについて研究をまとめ、国際会議で発表を行った。また、システムは公開し、メンテナンスを行った。コーパスデータの拡充については、肺疾患に関するコーパスの構築について研究成果をまとめ、論文誌で発表した。

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について、公開したデータキュレーションシステムのメンテナンスを行い、サービス提供を継続した。

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について、構築したウェブアプリケーションについてメンテナンス、サービス継続、および成果発表を行った。

③ コーパスデータの拡充：

これまで構築した肺疾患に関するコーパスの、アノテーションガイドライン、アノテーションスキーム及びコーパス自体について、研究成果をまとめ発表した。

(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は、計画通りアプリケーションのサービスを継続した。疾患ネットワーク構築技術については、計画通り国際会議での成果発表とウェブアプリケーションのサービス運用を行なった。ウェブアプリケーションは、文献を入力することで、そこから2次元および3次元の疾患ネットワークを自動的に構築するものであり、論文は、Association for Computational Linguisticsの会議のシステムデモンストレーション部門に採択された。また、肺疾患に関するコーパスについての学術論文は、Scientific Reports論文誌で採択された。

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) Mohammad Golam Sohrab, Khoa Duong, Goran Topić, Masami Ikeda, Nozomi Nagano, Yayoi Natsume-Kitatani, Masakata Kuroda, Mari Itoh, Hiroya Takamura, “Disease Network Constructor: a Pathway Extraction and Visualization”, Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations), pages 549-557, 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.

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