

## 医療知識基盤にもとづく高度医療情報利活用に関する研究

研究代表者 大江和彦 東京大学医学部附属病院企画情報運営部 教授

### 研究要旨

〔目的〕 医療知識基盤データベース（以下 MKB）を活用して、①医療安全支援機能、②電子カルテ DB やレセプト DB（NDB や DPC）（以下、既存 DB）の研究利用のための高品質データ抽出機能、の2つを実現する方法を検討する。またそのために基盤として必要となる日本語医療言語リソースを集約統合して利用できる基盤の整備を行い、これにより今後の医用人工知能開発の基盤となることを目指す。

〔方法〕 医療安全支援機能として、主要な医薬品について禁忌、投与注意、効能記述における疾患、病態、症状記述を抽出し、MKB に記述されている異常状態表現と対応付ける方法を開発し、MKB の異常状態表現を探索することで、疾患や検査結果、医薬品名そのものではなく、研究者が必要とする情報粒度でのデータ抽出を行う検索プログラムを開発する。

〔結果と結論〕 医療自然言語処理リソースを集約し医用文書にメタ情報を付与できるツールである医学用語自動アノテーションシステムを開発し、これを研究支援機能の開発と共通で利用できるようにした。また複数の言語リソースにまたがり横断的に意味上の集約を行うためのシソーラス検索 API である JMed-LexGrid を開発した。これらは国内 17 の言語リソースを集約した、医療自然言語処理を必要とする研究者やアプリケーションに非常に有用な言語リソースサービスとして提供できる。言語リソースを集約した、医療自然言語処理を必要とする研究者やアプリケーションに非常に有用な言語リソースサービスが開発され、医療安全支援機能を持った電子カルテや、電子カルテの文章データの二次利用システムに役立つと考えられる。

### 研究分担者

今井 健	東京大学大学院医学系研究科 准教授
河添悦昌	東京大学医学部附属病院 講師
古崎晃司	大阪大学産業科学研究所 准教授

### <研究協力者>

嶋本 公德	東京大学医学部附属病院
篠原 恵美子	東京大学医学部附属病院
Ma, Xiaojun	東京大学大学院

## A. 研究目的

医療知識基盤データベース（以下 MKB）は、疾患と異常状態（病態異常等）の因果関係により約 6000 疾患の定義が記述されたものである。本研究では①医療安全支援機能、②電子カルテ DB やレセプト DB（NDB や DPC）（以下、既存 DB）の研究利用のための高品質データ抽出機能、の 2 つを MKB を活用する方法を検討する。またそのために基盤として必要となる日本語医療言語リソースを集約統合して利用できる基盤の整備を行い、これにより今後の医用人工知能開発の基盤となることを目指す。

## B. 研究方法

医療安全支援機能として、処方オーダー時に警告や注意喚起を表示するシステムを試験開発できるようにするため、主要な医薬品について禁忌、投与注意、効能記述における疾患、病態、症状記述（以下、医薬品関連異常状態等）を抽出し、MKB に記述されている異常状態表現と対応付ける方法を開発する。また研究支援機能としては、MKB の異常状態表現を探索することで、疾患や検査結果、医薬品名そのものではなく、研究者が必要とする情報粒度でのデータ抽出を行う検索プログラムを開発する。初年度は医療自然言語処理リソースを集約し医用文書にメタ情報を付与できるツールを開発し、医薬品添付文書等に適用する。全体のイメージを図 1 に示す。

（倫理面への配慮）

本研究は情報システムの開発研究とデータ作成であり特別な倫理上の配慮は不要である。

## C. 研究結果と考察

医学用語自動アノテーションシステムの開発：

本研究だけでなく今後の多様な利用も可能となるよう、医療自然言語処理リソースを集約し医用文書にメタ情報を付与できるツールを開発し、これを研究支援機能の開発と共通で利用できるようにした。これを用いた医療安全支援機能と研究支援機能を設計し、一部のシステムモジュールを開発した。

具体的には、これまでの研究知見で得た言語リソースを標準病名、病名索引用語、病名修飾語、LiLak データベース、薬剤 JAPIC コード、HOT マスター、WEB 収集語、JLAC10 用語、診療行為マスター、手術処置マスター、解剖学用語、MedDRA、日本医学会用語 V 3 の全用語をひとつの Web リソースにまとめ、前記添付文書言語情報を文字列一致、形態素解析により自動的にアノテーションする汎用システム「医学用語自動アノテーションシステム」を開発した。これを用いて医薬品関連異常状態等を MKB の異常状態と対応づけられる用語を医薬品添付文書から抽出できる。また研究者が抽出したい情報粒度のやや大きい臨床概念を既存研究論文と研究計画書から収集できることを確認した（図 2）。

詳細は平成 28 年度および 29 年度の各総括報告書を参照されたい。

言語アノテーション情報に対する意味的集約のためのシソーラス検索 API:

#### JMed-LexGrid の開発

上述のように主要な言語辞書リソース 17 を用い、自然言語文リソースに対して文字列一致、形態素解析などにより汎用的かつ自動的にアノテーションする汎用システムの開発を行ったが、このシステムにてアノテーションされた言語素片タグに対し、複数の言語リソースにまたがり横断的に意味上の集約を行うためのシソーラス検索 API である JMed-LexGrid を開発した。

ここでは、全ての言語辞書リソースに共通した要素定義とメソッドを定義した。

`term2conceptIDs` メソッドにより、用語文字列からその用語が存在する特定リソース内のテーブル名とテーブル内の ID のセットを得ることができる。

また `conceptID2attributes` メソッドにより、特定リソース内テーブルと ID を指定することで、(1) 他の標準コードセットとの対応 (例: ICD10:2003)、(2)上位語の概念 ID セット、(3)下位語の概念 ID セット、(4)リソース内で定義された意味カテゴリ、(5)代表表記、(6)全同義語セット、(7)推論された意味カテゴリ、(8)その他の属性といった情報を取得することができる。

さらに、これらのメソッドを利用し上位・下位語や意味カテゴリ、ID コード、その他の情報を簡易的に取得するラッパーを用意し、JSON を用いた REST API として実装した。

このような API を用いて複数の言語辞書リソースを横断的に検索し、上位/下位概念や意味カテゴリの再帰的な探索や同義関係の取得、代表表記への名寄せを行うことで、

自然言語文に対する自動言語アノテーション情報に対してユーザーが求める粒度で意味的集約を行うことが可能である。

詳細は平成 29 年度総括報告書を参照されたい。

#### D. 結論

開発された JMed-LexGrid と医学用語自動アノテーションシステムは、国内 17 の言語リソースを集約した、医療自然言語処理を必要とする研究者やアプリケーションに非常に有用な言語リソースサービスとして提供できる。言語リソースを集約した、医療自然言語処理を必要とする研究者やアプリケーションに非常に有用な言語リソースサービスが開発され、医療安全支援機能を持った電子カルテや、電子カルテの文章データの二次利用システムに役立つと考えられる。

#### E. 健康危険情報

なし

#### F. 研究発表

##### 1. 論文発表

- 1) Kagawa R, Kawazoe Y, Ida Y, Shinohara E, Tanaka K, Imai T, Ohe K.: Development of Type 2 Diabetes Mellitus Phenotyping Framework Using Expert Knowledge and Machine Learning Approach. J Diabetes Sci Technol. 2016 Dec 7
- 2) Takeshi Imai, Emiko Shinohara, Masayuki Kajino, Ryota Sakurai, Kazuhiko Ohe, Kouji Kozaki, Riichiro Mizoguchi. An Ontological Framework for Representing Topological

- Information in Human Anatomy. In Proc. of the 5th International Conference on Biomedical Ontology (ICBO-BioCreative 2016), CEUR-ws.org Vol. 1747, IT506, Corvallis, Oregon, USA, Aug. 1-4, 2016.
- 3) Kawazoe Y, Imai T, Ohe K. :A Querying Method over RDF-ized Health Level Seven v2.5 Messages Using Life Science Knowledge Resources, JMIR Med Inform 2016;4(2)
  - 4) 山縣友紀, 古崎晃司, 今井 健, 大江和彦, 溝口理一郎. 疾患知識統合に向けた異常状態オントロジーの Linked Data 化. 人工知能学会論文誌 31(1):pp.LOD-A\_1-15, 2016.
  - 5) 篠原恵美子, 今井 健, 大江 和彦, 身体部位表現と解剖オントロジーのマッピングに関する基礎的検討, 医療情報学, 35(6): 275-282, 2016.
  - 6) Kagawa R, Kawazoe Y, Shinohara E, Imai T, Ohe K. The Impact of “Possible Patients” on Phenotyping Algorithms: Electronic Phenotype Algorithms Can Only Be Reproduced by Sharing Detailed Annotation Criteria. Stud Health Technol Inform. 2017;245:432-436.
  - 7) Ma X, Imai T, Shinohara E, Sakurai R, Kozaki K, Ohe K. A Semi-Automatic Framework to Identify Abnormal States in EHR Narratives. Stud Health Technol Inform. 2017;245:910-914
  - 8) Iwai S, Kawazoe Y, Imai T, Ohe K. Effects of Implementing a Tree Model of Diagnosis into a Bayesian Diagnostic Inference System. Stud Health Technol Inform. 2017;245:882-886.
  - 9) Kozaki K, Yamagata Y, Mizoguchi R, Imai T, Ohe K. Disease Compass- a navigation system for disease knowledge based on ontology and linked data techniques. J Biomed Semantics. 2017 Jun 19;8(1):22. doi: 10.1186/s13326-017-0132-2.
  - 10) 香川璃奈, 河添悦昌, 篠原恵美子, 今井健, 大江和彦, 疾患横断的な e-Phenotyping 手法開発を目的とした各疾患の特徴の検討, 医療情報学 37(Suppl.), 2017, 755-759.
  - 11) 河添悦昌, 香川璃奈, 今井健, 大江和彦, 診療情報による Phenotyping の現状・限界, 医療情報学 37(Suppl.), 2017, 169-172.
  - 12) 大江和彦, 医療における人工知能の活用と将来展望, 日本腎臓学会誌, 59 巻 7 号, 1060-1063(2017.10)
  - 13) 大江和彦, これからの医療における AI の活用と課題, 医薬品情報学 19 巻 3 号, N1-N3(2017.11)
  - 14) 倉沢央, 大江和彦, 糖尿病診療における人工知能の活用と展望, 糖尿病診療マスター 15 巻 12 号, 1054-1059(2017.12)
  - 15) 大江和彦, 新しい医療技術 医療における ICT 活用のイノベーション, 整形・災害外科(0387-4095)61 巻 1 号, 91-95(2018.01)
  - 16) 河添悦昌, 大江和彦, AI と ICT が変える医療, 腎臓内科・泌尿器科, 7 巻 2 号, 183-187(2018.02)

- 17) Ma X, Shinohara E, Han H, Ishii M, Imai T, Ohe K. Extracting Information on Lifestyle Issues from Clinical Narratives in EHR. 医療情報学 37(6), 313-321, 2018.
2. 学会発表
- 1) 香川 璃奈, 河添 悦昌, 篠原 恵美子, 今井 健, 大江 和彦, 高血圧の phenotyping 手法の開発および他疾患との比較検討, 医療情報学連合大会論文集 36 回 2 号, 770-773(2016.11).
- 2) 篠原 恵美子, Xiaojun Ma, 韓 浩, 石井 雅通, 今井 健, 大江 和彦, 医学知識と言語知識に基づく診療録テキストの深い解析手法の提案, 医療情報学連合大会論文集 36 回 2 号, 648-651(2016.11).
- 3) Xiaojun Ma, 篠原 恵美子, 韓 浩, 石井 雅通, 今井 健, 大江 和彦, EHR におけるナラティブの臨床テキストからの、ライフスタイル情報の抽出 (Extracting lifestyle information from narrative clinical texts in EHR)(英語), 医療情報学連合大会論文集 36 回 2 号, 642-646(2016.11).
- 4) 河添 悦昌, 香川 璃奈, 山口 亮平, 桜井 亮太, 篠原 恵美子, 大江 和彦, 電子的診療情報からの高次元特徴デー
- タを用いた EHR Phenotyping アルゴリズムの開発, 医療情報学連合大会論文集 36 回 1 号, 426-430(2016.11).
- 5) Xiaojun Ma, 篠原恵美子, 桜井亮太, 古崎晃司, 今井 健, 大江和彦. High throughput identification of patients' status from EHR. 第 2 回医療情報学会・人工知能学会 AIM 合同研究会 (東京), 医療情報学会・人工知能学会 AIM 合同研究会資料 SIG-AIMED-002-09, (2016.11)
- 1] 三浦祥治, 杉山博一, 赤堀光希, 高野良治, Xiaojun Ma, 篠原恵美子, 韓浩, 土井俊祐, 今井健, 大江和彦. 意味解析技術を用いた診療ガイドラインからの推奨ルール抽出. 第 4 回医療情報学会・人工知能学会 AIM 合同研究会 (東京), 医療情報学会・人工知能学会 AIM 合同研究会資料 SIG-AIMED-004-09, 2017.
- G. 知的財産権の出願・登録状況
1. 特許取得  
なし
2. 実用新案登録  
なし
3. その他  
なし