

厚生労働行政推進調査事業費補助金
(地域医療基盤開発推進研究事業)
総括研究報告書

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。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、医薬品添付文書や研究論文と研究計画書に適用する。

結果：主要な言語辞書リソース17について一つの Web リソースとしてまとめ、自然言語文リソースに対して文字列一致、形態素解析など汎用的に自動的にアノテーションする汎用システムを開発した。これを用いて医薬品添付文書から抽出された医薬品関連異常状態等を抽出できるようになった。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ようになった。医療知識基盤データベースの病態異常等の因果関係と既存の言語リソースを統合的に活用することが

できる

<研究分担者>

今井 健 東京大学大学院医学系研究科 疾患生命工学センター 准教授
河添悦昌 東京大学医学部附属病院企画情報運営部 講師
古崎晃司 大阪大学産業科学研究所 知識科学研究分野 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

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 を活用することで実現する。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

2 年計画で本研究目的の 2 本の柱である医療安全支援機能の実現、研究支援機能実現のための均質的で構造的なデータベース生成手法を開発し提案する。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（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、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し評価する。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、医薬品添付文書や研究論文と研究計画書に適用する。

C. 研究結果

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 V3 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。

表 1（末尾）表 1 に示すように、本 Web システムに集約した言語リソースはこれまでに 17 リソース、重複のある用語コード数は約 141 万エントリーに達した。また、図 1 に医学用語自動アノテーションシステムの言語リソース環境設定画面を示す。さらに図 2 に本システムの利用画面を示す。表 2 に、医薬品添付文書の警告、一般的注意、慎重投与などの記載を処理した例を示す。また、表 3 に臨床医学オントロジーに記述される複雑な表現の異常状態の処理例を示した。

D. 考察

医療言語リソースを利用した医療言語データの汎用自動アノテーションシステム

の開発により効率的に本研究を進めることができると考えられた。さらに MKB に記述される病態記述を探索する機能をこのシステムに追加することにより当初の研究目的を達成できると考えられる。

E. 結論

医療知識基盤データベースの病態異常等の因果関係と既存の言語リソース、および医薬品添付文書情報の処理結果を統合的に活用することにより医療安全支援機能と既存医療ビッグデータの研究利用支援の情報基盤を構築できるであろうことが示された。

F. 健康危険情報

特になし。

G. 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. pii: 1932296816681584. [Epub ahead of print]
- 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.

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(1347-8508)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(1347-8508)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

- 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 (1347-8508)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(1347-8508)36 回 1 号, 426-430(2016.11).
- 5) 馬 驍琚, 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)

H. 知的財産権の出願・登録状況
なし

表1 集約した言語（辞書）リソース一覧

用語集名	用語集内のリソース名	用語コード総数
略語集		
	カルテ&レセプト略語	15,064
	心エコー検査に頻出する英単語	324
ICD		
	見出し	11,082
医薬品 HOT コードマスター2017 年1月31日版		
	HOT13	51,382
看護実践用語標準マスター		
	看護観察編 ver. 3.1	5,674
臨床検査マスター		
	診療行為マスター	7,163
	分析物	3,219
	結果識別（固有）	3,484
	結果識別（共通）	99
	識別	1,543
	測定法	342
	材料	221
手術・処置マスター		
	手術・処置名称テーブル	10,344
症状・所見マスター＜身体所見編＞		
	症状用語	3,547
MEID 辞書		
	辞書語	229,952
技術用語辞書 PATRO		
	技術一般語	246,880

	医薬・バイオ	227, 288
解剖学用語		
	解剖学用語	8, 436
日本医学会医学用語辞典		
	日本医学会医学用語辞典	76, 608
T 辞書		
	医薬同義 T 辞書	100, 765
JAPIC 薬剤データベース		
	薬価と後発品	21, 519
	info_all	21, 518
	japic_who_atc	22, 326
LiLak2		
	病名	27, 838
	症状	3, 286
	部位	1, 869
MedDRA/J		
	1_hlgt_pref_term	334
	1_hlt_pref_term	1, 720
	1_low_level_term	73, 221
	1_low_level_term_j	66, 829
	1_pref_term	20, 808
ICD10 対応標準病名マスターV4		
	索引テーブル	102, 557
	修飾語テーブル	2, 269
	病名基本テーブル	25, 293
歯科病名マスター		
	歯科病名基本テーブル	3, 036
	歯科索引テーブル	13, 197
合計		1, 411, 037

図1 医学用語自動アノテーションシステムで使用する言語リソースの選択画面

略語集 (ver.0.1) ☒ カルテ&レセプト略語16000 (15064コード) (9662語) ☒ 心エコー検査に頻出する英単語 (324コード) (292語)	ICD (ver.10) ☒ isa略語 (0コード) (0語) ☒ 見出し (11082コード) (11057語)	医療機器データベース (ver.20170408) ☒ 全件一括 (0コード) (0語)	医薬品HOTコードマスター (ver.2017年1月) ☒ HOT13 (51382コード) (19599語) ☒ HOT9 (0コード) (0語) ☒ オプション (0コード) (0語)
看護実践用語標準マスター (ver.20161209) ☒ 看護観察編ver.3.1 (5674コード) (4778語) ☒ 看護観察編_結果 (0コード) (0語) ☒ 看護観察編_同義語 (0コード) (0語) ☒ 看護観察編_単位 (0コード) (0語) ☒ 看護行為編ver.3.1 (0コード) (0語)	臨床検査マスター (ver.256) ☒ 診療行為マスター (7163コード) (6950語) ☒ 分析物 (3219コード) (6708語) ☒ 17桁コード表 (0コード) (0語) ☒ 結果識別 (固有) (3484コード) (2681語) ☒ 結果識別 (共通) (99コード) (28語) ☒ 識別 (1543コード) (2890語) ☒ 測定法 (342コード) (653語) ☒ 材料 (221コード) (433語)	手術・処置マスター (ver.20161208) ☒ 手術・処置名称テーブル (10344コード) (13689語) ☒ レセ電・9CMテーブル (0コード) (0語) ☒ 患側テーブル (3コード) (3語) ☒ レセ電・Kコードテーブル (0コード) (0語) ☒ 性により制限される名称テーブル (0コード) (0語)	
症状・所見マスター<身体所見編> (ver.beta) ☒ 全身 (0コード) (0語) ☒ 境界 (0コード) (0語) ☒ 固定性 (4コード) (4語) ☒ 硬度 (4コード) (4語) ☒ リンパ節位置 (0コード) (0語) ☒ PHYXAM-beta (3547コード) (2199語) ☒ 腫瘍形態 (6コード) (6語) ☒ 大きさ (14コード) (14語) ☒ 表面性状 (4コード) (4語)	MEID辞書 (ver.2005頃入手) ☒ prolog (22952コード) (228195語)	形態素解析用辞書 (ver.20170425) ☒ ComeJisyoV5-1 (0コード) (0語) ☒ mecab_ipadic-2.7.0 (0コード) (0語) ☒ mecab_jumandic-7.0 (0コード) (0語) ☒ naist-jdic-0.4.3 (0コード) (0語) ☒ unidic-2.1.2 (0コード) (0語)	技術用語辞書PATRO (ver.20170412大江先生) ☒ 技術一般語 (246880コード) (223769語) ☒ 医薬・バイオ (227288コード) (211699語)
解剖学用語 (ver.13) ☒ 解剖学用語 (8436コード) (7495語)	日本医学会医学用語辞典 (ver.V3当初(部内限定)) ☒ 日本医学会医学用語辞典 (76608コード) (68627語)	T辞書 (ver.2016年11月版) ☒ 医薬同義T辞書 (100765コード) (99839語)	
umls (ver.2016AB) ☒ MRCONSO(concept names and sources) (0コード) (0語)	JAPIC薬剤データベース (ver.2016年05月11日) ☒ 薬価と後発品 (21519コード) (18845語) ☒ info_all (21518コード) (24994語) ☒ japic_who_atc (22326コード) (1153語)	LilLak (ver.2) ☒ 病名親子関係 (0コード) (0語) ☒ 病名 (27838コード) (27834語) ☒ 症状親子関係 (0コード) (0語) ☒ 症状 (3286コード) (5149語) ☒ 部位親子関係 (0コード) (0語) ☒ 部位 (1869コード) (1868語)	
東大病院ローカルマスター (ver.20170412) ☒ 検査項目 (0コード) (0語)	MedDRA/J (ver.17.1) ☒ l_hlgt_pref_term (334コード) (334語) ☒ l_hlgt_hlt_comp (0コード) (0語) ☒ l_hlgt_pref_term_j (334コード) (334語) ☒ l_hlgt_pref_term (1720コード) (1720語) ☒ l_hlt_pref_term_j (1720コード) (1717語) ☒ l_hlt_pref_comp (0コード) (0語) ☒ l_soc_intl_order (0コード) (0語) ☒ l_low_level_term (73221コード) (73218語) ☒ l_low_level_term_j (66829コード) (52838語) ☒ l_md_hierarchy (0コード) (0語) ☒ meddra_history_japanese_english (0コード) (0語) ☒ meddra_release (0コード) (0語) ☒ l_pref_term (20808コード) (20808語) ☒ l_pref_term_j (20808コード) (20808語) ☒ l_smq_content (0コード) (0語) ☒ l_smq_list (0コード) (0語) ☒ l_smq_list_j (0コード) (0語) ☒ l_soc_term (26コード) (26語) ☒ l_soc_hlgt_comp (0コード) (0語)	ICD10対応標準病名マスター (ver.400) ☒ 索引テーブル (102557コード) (99903語) ☒ 修飾語テーブル (2209コード) (2209語) ☒ 病名基本テーブル (25293コード) (25293語)	
歯科病名マスター (ver.200) ☒ 歯科病名基本テーブル (3036コード) (3035語) ☒ 歯科索引テーブル (13197コード) (12944語)	歯科手術・処置マスター (ver.B版) ☒ 歯科手術・処置マスター (0コード) (0語)		

図2 医学用語自動アノテーションシステムの使用例

医学用語の自動アノテーション

設定

・ビタミンC欠乏症の予防および治療(壊血病、メルレル・バロー病)12. ビタミンCの需要が増大し、食事からの摂取が不十分な際の補給(消耗性疾患、妊娠婦、授乳婦、はげしい肉体労働時など)13. 下記疾患のうち、ビタミンCの欠乏または代謝障害が関与すると推定される場合!(1). 毛細管出血(鼻出血、歯肉出血、血尿など)!(2). 薬物中毒!(3). 副腎皮質機能障害!(4). 骨折時の骨基質形成・骨癒合促進!(5). 肝斑・雀卵斑・炎症後の色素沈着!(6). 光線過敏性皮膚炎!なお3. の効能に対して効果がないのに月余にわたって漫然と使用すべきでない。

松

竹

梅

・ビタミンC欠乏症の予防および治療(壊血病、メルレル・バロー病)12. ビタミンCの需要が増大し、食事からの摂取が不十分な際の補給(消耗性疾患、妊娠婦、授乳婦、はげしい肉体労働時など)! 3. 下記疾患のうち、ビタミンCの欠乏または代謝障害が関与すると推定される場合!(1). 毛細管出血(鼻出血、歯肉出血、血尿など)!(2). 薬物中毒!(3). 副腎皮質機能障害!(4). 骨折時の骨基質形成・骨癒合促進!(5). 肝斑・雀卵斑・炎症後の色素沈着!(6). 光線過敏性皮膚炎!なお3. の効能に対して効果がないのに月余にわたって漫然と使用すべきでない。

用語	リソース	テーブル	代表表記	コード体系	コード
ビタミン	MEID辞書	prolog	ビタミン ; ビタミン類		
ビタミン	技術用語辞書PATRO	技術一般語	ビタミン	PATRO	K01\00602770
ビタミン	技術用語辞書PATRO	医薬・バイオ	ビタミン	PATRO	K09\00384290
ビタミン	日本医学会医学用語辞典	日本医学会医学用語辞典	ビタミン		
ビタミン	LiLak	症状	ビタミン	LiLak	
ビタミンC	臨床検査マスター	診療行為マスター	ビタミンC		
ビタミンC	臨床検査マスター	分析物	ビタミンC	JLAC10_bun	3G060
ビタミンC	MEID辞書	prolog	ビタミンC		
ビタミンC	技術用語辞書PATRO	技術一般語	ビタミンC	PATRO	K01\00846830
ビタミンC	技術用語辞書PATRO	医薬・バイオ	ビタミンC	PATRO	K09\00274270
ビタミンC	日本医学会医学用語辞典	日本医学会医学用語辞典	ビタミンC		
ビタミンC	T辞書	医薬同義T辞書	ビタミンC	Tdic	E00125
ビタミンC	JAPIC薬剤データベース	info_all	アスコルビン酸	HOT9	I07120704
ビタミンC	JAPIC薬剤データベース	info_all	アスコルビン酸	JAPIC_tenpu	00054584
ビタミンC	JAPIC薬剤データベース	info_all	アスコルビン酸	JAPIC_tenpu_branch	002
ビタミンC	JAPIC薬剤データベース	info_all	アスコルビン酸	YJ	3140001B2100

表 2 医学用語自動アノテーションシステムの使用による医薬品添付文書の分析例

記載 区分	医薬品添付文書の処理事例
警告	.ビタミンC欠乏症の予防および治療(壊血病,メルレル・パロー病)!2.ビタミンCの需要が増大し,食事からの摂取が不十分な際の補給(消耗性痔,妊産婦,授乳婦,はげしい肉休労働時など)!3.下記疾患のうち,ビタミンCの欠乏または代謝障害が関与すると推定される場合!(1).毛細管出血(鼻血,歯肉出血,血尿など)!(2).薬物中毒!(3).副腎皮質機能障害!(4).骨折時の骨基質形成・骨癒合促進!(5).肝斑・雀斑・炎症後の色素沈着!(6).光過敏性皮膚炎!なお3.の効能に対して効果がないのに月余にわたって漫然と使用すべきでない。
一般的 注意	作用発現状況の概要!本剤は使用成績調査等の副作用発現頻度が明確となる調査を実施していない。!その他の副作用!胃腸:悪心,嘔吐,下痢(例不明)等があらわれることがある。!
警告	基本的注意!1.サリチル酸系製剤の使用実態は我が国と異なるものの,米国においてサリチル酸系製剤とライ症候群との関連性を示す疫学調査報告があるので,本剤を15歳未満の水痘,インフルエンザの患者に投与しないことを原則とするが,やむを得ず投与する場合には,慎重に投与し,投与後の患者の状態を十分に観察すること。!ライ症候群:小児において極めてまれに水痘,インフルエンザ等のウイルス性疾患の先行後,激しい嘔吐,意識障害,痙攣(急性脳浮腫)と肝臓ほか諸臓器の脂肪沈着,ミトコンドリア変形,AST(GOT)・ALT(GPT)・LDH・CK(CPK)の急激な上昇,高アンモニア血症,低プロトロンビン血症,低血糖等の症状が短期間に発現する高死亡率の病態である。!2.解熱鎮痛剤による治療は原因療法ではなく対症療法であることに留意すること。!3.慢性疾患(関節リウマチ,変形性関節症等)に対し本剤を用いる場合には,次の事項を考慮すること。!(1).長期投与する場合には定期的に臨床検査(尿検査,血液検査及び肝機能検査等)を行うこと。!また,異常が認められた場合には減量,休業等の適切な措置を講ずること。!(2).薬物療法以外の療法も考慮すること。!4.急性疾患に対し本剤を用いる場合には,次の事項を考慮すること。!(1).疼痛,発熱の程度を考慮し投与すること。!(2).原則として同一の薬剤の長期投与を避けること。!(3).原因療法があればこれを行うこと。!5.患者の状態を十分観察し,副作用の発現に留意すること。!過度の体温下降,虚脱,四肢冷却等があらわれることがあるので,特に高熱を伴う小児及び高齢者又は消耗性疾患の患者においては,投与後の患者の状態に十分注意すること。!6.感染症を不顕性化するおそれがあるので,感染による炎症に対して用いる場合には適切な抗菌剤を併用し,観察を十分に行い慎重に投与すること。!7.他の消炎鎮痛剤との併用を避けることが望ましい。(「3.相互作用」の項参照)!8.高齢者及び小児には副作用の発現に特に注意し,必要最小限の使用にとどめるなど慎重に投与すること。!(「5.高齢者への投与」,「7.小児等への投与」の項参照)!9.手術前1週間以内にアスピリンを投与した例では失血量が有意に増加したとの報告があるので,術前の投与は慎重に行うこと。!
禁忌	禁忌(次の患者には投与しないこと)!1.本剤又はサリチル酸系製剤に対し過敏症の既往歴のある患者!2.消化性潰瘍のある患者[胃出血の発現又は潰瘍が悪化するおそれがある。](ただし,「1.慎重投与」の項参照)!3.重篤な血液の異常のある患者[血液の異常をさらに悪化させるおそれがある。](「4.副作用」の項参照)!4.重篤な肝障害のある患者[肝障害をさらに悪化させるおそれがある。]!5.重篤な腎障害のある患者[腎障害をさらに悪化させるおそれがある。]!6.重篤な心機能不全のある患者[心機能をさらに悪化させるおそれがある。]!7.アスピリン喘息(非ステロイド性消炎鎮痛剤等による喘息発作の誘発)又はその既往歴のある患者[重篤な喘息発作を誘発させるおそれがある。]!8.出産予定日12週以内の妊婦
慎重 投与	作用発現状況の概要!本品は使用成績調査等の副作用発現頻度が明確となる調査を実施していないため,発現頻度は不明である。!その他の副作用!(表開始)!発現部位等:頻度不明!その他:心悸亢進,のぼせ,舌のしびれ,悪心等!

表 3 臨床医学オントロジーに記述される複雑な表現の異常状態の処理事例

小腸の原始カイロミクロンあるいは肝臓の原始VLDLに脂質を転送するミクロソームトリグリセリド転送蛋白をコードする遺伝子の機能喪失型変異
膵臓の発生とインスリン遺伝子転写を調節する転写因子であるインスリンプロモーター因子(IPF-1)の変異に起因する糖尿病
血漿トリグリセリド(VLDL)とコレステロール(LDL)の中等度上昇とHDLコレステロールの低下