

臨床医学オントロジーにおける異常状態と診療情報との自動マッピングに向けた基礎的検討

A Fundamental Consideration for Automated Mapping between Medical Ontology and Clinical Information

今井 健*¹
Takeshi Imai

林 亜紀*²
Aki Hayashi

桜井亮太*²
Ryota Sakurai

馬 驍珺*¹
Xiaojun Ma

篠原 恵美子*²
Emiko Shinohara

古崎 晃司*³
Kouji Kozaki

大江 和彦*¹
Kazuhiko Ohe

*¹ 東京大学大学院医学系研究科
Graduate School of Medicine,
The University of Tokyo

*² 東京大学医学部附属病院
The University of Tokyo Hospital

*³ 大阪大学産業科学研究所
The Institute of Scientific and Industrial
Research (ISIR), Osaka University

Automated mapping between Medical Ontology and clinical information is indispensable for applying Medical Ontology to clinical decision support system. This study discusses the technical requirements to develop the automated mapping module between them, based on the analysis of 4,718 (642 distinct) abnormal states from 148 chronic diseases in the Japanese Medical Ontology. We found that 382 abnormal states (59.5%) are potentially be able to map to the combinations of clinical information, such as text information, laboratory data, prescription and treatment data using natural language processing and machine learning techniques.

1. 背景と目的

医療の場では、電子カルテの普及に伴い電子化された診療情報の蓄積が進んでいる。これを2次的に利用し、患者の診療の記録と参照に用いるだけでなく、安全な診療を実現するための注意喚起や、診療上の意思決定過程の確認、電子カルテ入力支援など、診療上の種々の支援に積極的に活用することが期待されている。これには様々な技術が必要であるが、共通する基盤的な知識データベースとして、医学用語概念同士の意味関係を記述した臨床医学オントロジーの開発は重要な課題である。このような背景のもと、これまで筆者らは厚生労働省からの受託研究により、本邦初の大規模臨床医学オントロジーの開発を遂行した(平成19~21年度, 22~24年度, 25~27年度)。そこでは疾患を異常状態の因果連鎖の総体として捉えて概念化を行っている[IMAI2009, 大江2010, MIZOGUCHI2011]。

この臨床医学オントロジーは医師の共通概念理解を記述したものであり、診療情報とは必ずしも一致せず、直接的に診療データベースに適用することはできない。例えば慢性腎臓病における「糸球体濾過量の低下」という異常状態概念は直接観測や記録はされず、診療情報として記録される尿検査・血液検査の特定の項目の数値から推定されるものである。従って、臨床医学オントロジーを診療データベースに適用し、上述の様々な目的に活用するためには、患者診療情報とオントロジー中の概念とを自動マッピングする手法の確立が必須である。

そこで本研究では、生活習慣病関連の慢性疾患領域を対象に、第一段階として、オントロジーに記述されている異常状態概念と電子カルテに記録されている個々の患者の診療情報(検査値, 文章記載, 画像・波形情報)との自動マッピングに必要な技術的課題, 可能な範囲について考察する。

連絡先: 今井 健, 東京大学大学院医学系研究科 疾患生命科学センター 医工情報研究領域, 〒113-8655 東京都文京区本郷 7-3-1, TEL:03-5841-3014, imai@m.u-tokyo.ac.jp

2. 材料と方法

2.1 臨床医学オントロジーにおける疾患と異常状態記述

図1に本研究で用いた臨床医学オントロジーにおける疾患記述と異常状態の関係について例を示す(2型糖尿病)。各疾患概念は複数の異常状態の因果連鎖として定義され、下位疾患では上位疾患の異常状態を特殊化するか、あるいは異常状態が追加されている。従って、1つの異常状態(“高血糖”など)は複数の疾患に現れることがある。各疾患に対しては、その定義となる「疾患定義連鎖」、また定義には含まれないがその疾患で見られる異常状態連鎖である「派生連鎖」が記述されている。定義連鎖中の異常状態はその疾患の患者全てに起こっていると考えられるが、派生連鎖中の異常状態はその限りではなく、各患者によって異なるものである。従って、派生連鎖まで含めた異常状態と診療情報をマッピングすることは、詳細に患者個別の病状を把握するために重要である。

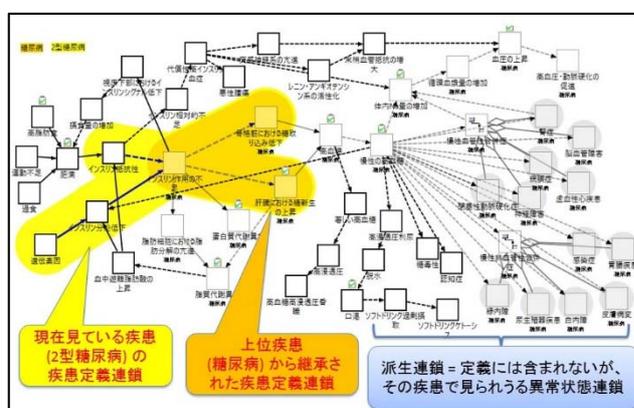


図1 疾患記述と異常状態の関係(2型糖尿病)

2.2 本研究のフレームワーク

図2に本研究全体のフレームワークを示す。診療情報には患者基本情報、病名、処方内容、カルテ文章、各種検体検査結果・画像検査結果などが含まれるが、これらを自然言語処理技術やルール記述あるいは機械学習を組み合わせたマッピングモジュールにより、図1で示した疾患連鎖中の異常状態と自動的にマッピングし、各種診断支援等に活用することが本研究全体の目的である。マッピング可能な異常状態には心疾患における“ST上昇”のように心電図中の構造化所見、あるいはテキスト表現と対応付け可能なもの、尿検査や血液検査結果の数値情報あるいはその複数の組み合わせから推定可能なもの、処方されている薬や施術内容から推定可能なものもある。しかしその一方で慢性腎臓病における“1 α -水酸化酵素の活性低下”のように機序として体内で起こっていると考えられているが、観測が困難なものもある。このようなものは疾患連鎖中では“missing node”となるが、可能なものだけをマッピングし、またそれを時系列可視化することにより、図2に示すように患者病状進展の視覚的把握にも応用可能であると考えられる。

本研究ではこのための第一段階として、各異常状態を分析することにより、どのような診療情報を用いることで、どの程度マッピング可能か見積もることとした。

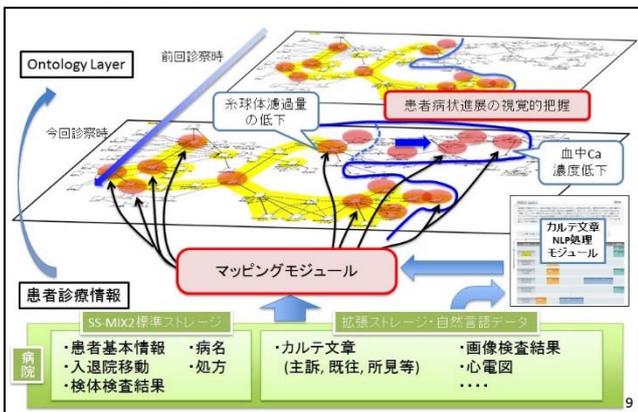


図2 本研究全体のフレームワーク（オントロジーと診療情報のマッピングによる患者病状進展の視覚的把握）

2.3 対象領域の選定

図2に示した通り、本研究では応用事例として患者病態進展の視覚的把握を想定している。そのため、異常状態の分析対象として主に生活習慣に関連の深い慢性疾患6領域(1)虚血性心疾患、(2)脳梗塞、(3)高血圧、(4)脂質代謝異常、(5)腎疾患(6)糖尿病を選定した。これらは病態の進行に伴い互いに他の領域の疾患へ移行する可能性があり、疾患同士の巨大な因果連鎖を形成するもので、本研究の目的に適している。またこの目的のため、(2)脳梗塞、(3)高血圧、(5)腎疾患の領域については、それぞれ他の領域からへ移行する可能性のある一部の疾患に限定した。例えば妊娠性高血圧症、薬剤性高血圧症などは上記に含まれないため、対象としていない。

2.4 異常状態の分析

次に、2.3で選定した6疾患領域につき、それぞれの最上位の疾患から辿られる全ての下位疾患を対象とし、全異常状態を抽出した。また東大病院の医師2名により、各異常状態が患者に起こっていると潜在的に推定できる可能性があるものについて、その判断材料のアノテーションを行った。

例えば、疾患「心筋梗塞」に含まれる異常状態「心室収縮力の低下」であれば、心エコー、MRI、心臓カテーテル左室造影の各検査所見中の当該テキスト記述が判断材料となる。また疾患「慢性腎臓病」の異常状態「腸管からのCa吸収低下」であれば、血液検査結果のCa値の低下、PTH値の上昇、あるいはその患者に活性型ビタミンD製剤が処方されていること、が判断材料となる。ここでは単独あるいは複数の診療情報から推定できる可能性がある場合には「マッピング可能性あり」と定義し、具体的なアルゴリズムが想定できなくても構わないこととした。

また、対象とする情報の種類(テキスト表現/数値/コード)、あるいはそれが記載されている情報ソース(カルテ文章、検査結果値、画像診断レポート、処方データ等)により、後に自動マッピングで用いる技術が異なると考えられるため、各異常状態に対応づけられる診療情報を、(1)カルテ記載文章(病名・所見情報、テキスト)(2)血液・尿検査(定性・定量情報)(3)心電図検査(検査所見、テキスト・構造化データ)(4)画像検査(検査所見、主にテキスト)(5)治療内容(処方内容・施術内容、テキストもしくはコード)の5つのカテゴリに分類した。但し心・頸動脈エコー、心・脳カテーテル、核医学検査は全て画像検査に含めた。

3. 結果

3.1 異常状態分析結果

今回対象としたカテゴリに含まれる疾患(下位疾患含む)とこれらの疾患に含まれる異常状態概念数の一覧を表1に示す。6つの疾患カテゴリに含まれる148疾患中において、のべ4,718個の(異なり概念数642個)の異常状態概念が存在した。

表1 各疾患カテゴリと疾患・異常状態数

疾患カテゴリ	疾患数	のべ異常状態数	異なり異常状態数
虚血性心疾患	52	2,755	133
脳梗塞	3	30	22
高血圧	9	61	47
脂質代謝異常	39	233	150
腎疾患	3	121	118
糖尿病	42	1,518	172
計	148	4,718	642

次に、642個の異なり異常状態概念について、診療情報(あるいはその組み合わせ)からマッピングできる可能性があるものを、その用いる情報の区分に応じて(1)カルテ記載情報、(2)尿・血液検査、(3)心電図所見、(4)画像所見、(5)治療内容に区分した結果を表2に示す。

表2 異常状態のマッピング可能性と用いる情報区分

マッピング可能性あり		382
(1)カルテ記載文章	病名	85
	所見	95
(2)血液・尿検査	尿検査結果	24
	血液検査結果	113
(3)心電図検査	心電図検査所見	12
(4)画像検査	MRI・CT・エコー・カテーテル検査所見	89
(5)治療内容	処方内容	17
	施術内容	26
マッピング不可能		260
計		642

全異なり異常状態 642のうち、382個(59.5%)の異常状態は表 2 に示す診療情報からマッピングできる可能性があることが判明した。但し、判断材料が複数の情報の組み合わせであることもあり、重複が存在する。

4. 考察

4.1 マッピング可能な範囲

3.1 の結果の通り、382 個 (59.5%) の異なり異常状態は診療情報(及びその組み合わせ)から潜在的にマッピングできる可能性があることが判明した。このマッピングのためには少なくとも 4.2 で考察する複数の技術が必要と考えられるが、全てが可能となるとは限らない。技術的な精度の問題、並びに必要な情報が全て記載されていない可能性があるためである。その患者に起こっている異常状態であっても、現在問題となっている疾患(その診療科の医師が興味とする対象)以外の情報は記載されないことも多い。その意味でこれは本研究がマッピング対象とし得る異常状態の割合の上限である。

一方、臨床では疾患連鎖中のいくつかの「キーとなる」異常状態が重要な意味を持ち、医師は判断のために必ずしも全てのノードを必要としない。従ってマッピング可能な上限が約 6 割程度であっても慢性疾患の病状進展把握、あるいは観測情報から候補疾患を自動的に絞り込む鑑別診断支援の目的にとつては、十分であると筆者らは考えている。

4.2 異常状態概念との自動マッピングに必要な技術

(1) テキストからの情報抽出技術

診療情報における文章記載はカルテ文章のみならず各種検査結果レポートなど多岐にわたる。表 2 において、テキストからの情報抽出が必要な対象は、カルテ記載文章からの病名・所見抽出(“糖尿病”, “浮腫” など)、心電図検査・画像検査からの所見抽出(“ST 上昇”, “心室内血栓” など)、処方・施術内容(“Ca 拮抗薬投与”, “経皮的心室中隔穿孔閉鎖術” など)である。このうち、病名については患者の病名リストとして構造化されているが、検査行為の保険請求上必要なため付与されたいわゆる保険病名が記載されていることもあり、必ずしも患者の病態を反映していない。従ってカルテ文章からの抽出も補完的に必要であろう。処方内容については処方オーダーからそのコード情報を得ることができるが、他院での処方内容についてはカルテ文章に頼らざるを得ない。また施術内容についても包括医療費支払い制度の DPC データから得ることができるが、「〇〇摘出術後」など DPC データに含まれない患者状態については、テキスト表現から得る必要があると考えられる。

(2) 検査結果定量・定性情報からの異常状態推定技術

血液・尿検査結果は定量情報と異常状態の有無といった定性情報が含まれ、単独あるいは組み合わせにより、ある特定の異常状態を推定できる可能性がある。このうち「ある特定の数値が〇〇以上である」といった判断クライテリアが存在するものについてはルール処理によるマッピングが可能と考えられる。筆者らはこれまで、異常状態を対象・属性・値の組で構造化して記述した異常状態オントロジー[YAMAGATA 2014, 山縣 2014]を開発しているが、ルール処理の効率的な生成のために、この構造化記述を利用することが有用と考えられ、今後の課題である。

一方、複数の情報の組み合わせにて総合的に判断されるものについては、機械学習の利用が有効と考えられる。例えば慢性腎臓病における「機能するネフロン数の減少」という異常状態

概念は直接的には観測が困難で、「ネフロン」の「数」が「減少」すること、という構造的な意味記述は、診療情報とのマッピングには役に立たない。しかし代替の観測情報として、血液検査の尿素窒素値の上昇、クレアチニン値の上昇、血清 $\beta 2$ ミクログロブリン値の上昇から推定が可能と考えられる。今後これらの情報の特徴量として用いたマッピング手法の開発と共に、時系列変化(増加・減少など)の取り扱いが重要な課題である。

(3) 複数情報の統合、コンテキストの利用

上述の (1), (2) は単体で用いるだけではなく、組み合わせで用いることも考えられる。さらにその患者に特定の診断が既に付いているという条件(コンテキスト)を利用するなど、複数の情報を統合して推論する仕組みも重要であると考えられる。

例えば前述の“ 1α -水酸化酵素の活性低下”はマッピング不可能と判定されたが、その前後のノードである“糸球体濾過量の低下”、“活性型ビタミン D [1,25(OH)₂D] の産生低下”は、それぞれ「尿・血液検査結果」、「活性型ビタミン D 製剤の処方」から推定できる可能性がある。従って「慢性腎臓病」というコンテキストの下で前後のノードが起こっていれば、それに挟まれた異常状態も発生しているはずである、といった推論が特定の条件下で成立する可能性があり、今後さらなる精査を行う予定である。

5. 結語

本研究では、生活習慣病関連の慢性疾患領域を対象に、オントロジーに記述されている異常状態を推定するための診療情報(検査値、文章記載、画像・波形情報)の種別について分析し、自動マッピングの実現に必要な技術的課題、対象とできる範囲について考察した。

謝辞

本研究の一部は、日本医療研究開発機構(AMED)「臨床研究等 ICT 基盤構築研究事業(課題番号: 15lk1010003h0001)」並びに JSPS 科研費(課題番号: 26460857)により、助成を受けたものである。

参考文献

- [YAMAGATA 2014] Yamagata Y, Kozaki K, Imai T, Ohe K, Mizoguchi R. An ontological modeling approach for abnormal states and its application in the medical domain. *J Biomed Semantics*. 2014 May 21;5:23. doi: 10.1186/2041-1480-5-23. eCollection 2014.
- [山縣 2014] 山縣友紀, 古崎晃司, 今井 健, 大江和彦, 溝口理一郎. 疾患における異常状態オントロジーの構築. *医療情報学* 34(3), pp.101-117, 2014.
- [MIZOGUCHI 2011] Mizoguchi R, Kozaki K, Kou H, Yamagata Y, Imai T, Waki K, Ohe K. River Flow Model of Diseases. *Proc. of International Conference on Biomedical Ontology 2011 (ICBO2011)*, pp.63-70, 2011.
- [大江 2010] 大江和彦, 今井 健. 臨床医学知識処理を目指した医療オントロジー開発. *人工知能学会誌* 25(4):pp.493-500, 2010.
- [IMAI 2009] Imai T, Kou H, Zhou J, Kozaki K, Mizoguchi R, Ohe K. Japan Medical Ontology Development Project for Advanced Clinical Information Systems. *Proc. Of 10th International HL7 Interoperability Conference 2009 (IHIC2009)*, pp.42-46, 2009.