

一般口演

一般口演22

医療データ解析

2017年11月22日(水) 16:00 ~ 17:15 C会場 (10F 会議室1001)

[3-C-3-OP22-4] 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香川 璃奈¹, 河添 悦昌², 篠原 恵美子², 今井 健³, 大江 和彦¹ (1.東京大学大学院, 2.東京大学医学部附属病院, 3.東京大学疾患生命工学センター)

【背景】 電子的診療データを用いた解析で重要な情報の一つは病名であるが、保険目的に登録される病名は信頼性が低く、DPC病名の利用も入院患者に限定される。そこで複数種類の診療データから特定の疾患を有する症例を自動で抽出する e-phenotyping手法が必要とされる。

【目的と手法】 e-phenotyping手法開発の観点から各疾患について電子カルテデータの特徴を明らかにすることを目的とする。本稿では複数の疾患についてカルテに明示的にその疾患の有無が記載されるか否かを調査した。慢性疾患か急性疾患か、患者数の多寡、診断において重視される検査種、といった臨床的特徴の異なる10疾患を対象とした。医師2名が目視でカルテを確認し、対象疾患の有無を判断する上で必要とした情報の分類を記録した(カルテを確認した対象患者数は疾患毎に異なり280人-650人(平均461.9人))。用いた分類は対象疾患のうち4疾患の記録作業を通して筆者らが作成と改良を行い、残りの6疾患に対して同じ基準を用いた。その結果から、カルテ自然文における対象疾患名の記載や診断基準を満たすなどの明示的な情報のみで判断が可能な症例のみの疾患と、鑑別疾患や類似する疾患の記載、治療内容など、対象疾患については非明示的な情報が必要な症例が存在する疾患の調査を行った。

【結果】 対象疾患のうち2型糖尿病、クローン病、本態性高血圧では非明示的な情報を用いないと疾患の有無の判断が変わる症例が存在した。その他の7疾患のうち乳癌、腎細胞癌、肺塞栓について、文字列一致のルールで患者抽出を行ったところ、感度は95%以上、陽性的中率は80%以上であった。

【考察】 2タイプの疾患の違いの原因として、2型糖尿病に対する二次性糖尿病のように鑑別が難しい疾患の存在、患者が高齢のため厳密に鑑別疾患との区別を行わない症例の存在などが考えられた。このような疾患は、検査結果などの構造化データや自然文データなどの総合的な利用が e-phenotyping手法開発に必要と考えられる。

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

香川 璃奈*¹ 河添 悦昌*² 篠原 恵美子*² 今井 健*¹ 大江 和彦*¹

*¹ 東京大学医学系研究科 *² 東京大学医学部附属病院

Study of the features of EHR data for development of the disease-independent framework for e-phenotyping techniques

Kagawa Rina *¹ Kawazoe Yoshimasa *² Shinohara Emiko *² Imai Takeshi *¹

Ohe Kazuhiko *¹

*¹ Department of Biomedical Informatics, Graduate School of Medicine, The University of Tokyo, *² Department of Healthcare Information Management, The University of Tokyo Hospital

e-phenotyping is an automated technique for identifying patients who have a particular disease based on electronic health records (EHRs). A framework for developing new e-phenotyping methods, which is independent of existing e-phenotyping methods or specific diseases, is necessary. However, there is no data that can be the basis of the framework. Therefore, first, we studied whether medical doctors use only explicit information or use both of explicit and implicit information for manual annotation of whether one patient has a particular disease or not. Based on this result, we assumed a classifier that can correctly judge whether a patient has a particular disease with only explicit information. Then, we calculated the maximum value of recall and precision when applying the classifier as an e-phenotyping method. To extract negative patients, our results showed that 97% or more sensitivity or precision could be achieved via e-phenotyping method with only explicit information. While, only 95.2% sensitivity or 81.3% precision could be achieved at least. The expected improvement of sensitivity is only 1.7% at the maximum when the data, which are not easy to handle electronically currently, can be easily handled. This result suggests that theoretical sensitivity can be achieved if technical improvement can be done sufficiently.

Keywords: e-phenotyping, EHR, SS-MIX2

1. 背景

電子的診療録(以後、EHR データと書く)を用いた解析で重要な情報の一つは病名であるが、保険目的に登録される病名は信頼性が低く、DPC 病名の利用も入院患者に限定される。そこで複数種類の EHR データから特定の疾患を有する症例を自動で抽出する e-phenotyping 手法が必要とされる[1]。本研究で議論する e-phenotyping 手法においては、既往歴含める対象疾患の有無の判定のみを対象とし、発症時期や罹患期間などの同定は対象外とする。

ここで e-phenotyping 手法の開発の際には、下記の 3 段階のプロセスを経る。

- (1)対象患者に正解データ(対象疾患の有無)を付与する。この作業を annotation と呼ぶ。
- (2)e-phenotyping 手法を開発する。
- (3)(1)と(2)を用いて精度の評価を行う。

ここで、(1)のプロセスの詳細が明らかであれば、(2)は(1)を再現するように作成すれば良い。(1)を専門家に任せるブラックボックスにしたままで(2)の手法開発を行うと下記のような問題点を克服できない。

まず、(2)で選択した手法や利用した項目が適切なのかどうか、評価から得られた精度の数値からしか判断を行えない。たとえば、仮にカルテ自然文における対象疾患名の記載や診断基準を満たすなどの明示的な情報のみで判断が可能な疾患が存在するならば、表記ゆれや英語表記なども含めて対象疾患名を正しく抽出できる自然言語処理技術の開発や、ルールの厳密な記載などが最優先課題だろう。その一方、鑑別疾患や類似する疾患の記載(例:「糖尿病は、1型や二次性は考え難い」という記載から2型糖尿病の可能性が高いと推測できる)、治療内容、対象臓器についての言及(例:「腎臓に異常無し」という記載から、腎

細胞癌が存在しないだろうと推測できる)など、判断にコンテキストが必要な対象疾患に対しては、様々な EHR データを組み合わせた手法の開発などが優先されるだろう。このように(1)の詳細を明らかにできれば、(2)で選択した手法や利用した項目が適切なのかをデータや疾患の特徴に基づいて議論できる枠組みを作成でき、疾患さらには施設を超えた e-phenotyping 開発フレームワークの基盤となりうると考える。疾患依存性になるべく低い e-phenotyping 開発フレームワークを作ろうとしても、EHR データからボトムアップに作成するための基盤となる先行研究が調べた限りで存在しなかった。疾患依存性になるべく低い e-phenotyping 開発フレームワーク作成の試みの先行研究は、既存の e-phenotyping 手法に基づいたフレームワークの作成のみにとどまっている[4]。

また、仮に(1)で必ず用いられる情報がある場合、その電子的な利用が難しいことが原因で精度が低いのであれば、(2)での手法開発よりもその情報の電子化を進める方が優先されるべきである。

2. 目的

本研究では、疾患依存性になるべく低い e-phenotyping 開発フレームワーク作成のための計算機処理における課題や用いるべき技術の選択、また、整備が必要なデータの有無を検討する。その検討に用いる基盤的なデータとして、医師が EHR データを目視で確認して対象疾患の有無を判断する際に用いる情報を調査する。

具体的には、医師が用いる情報が、カルテ自然文における対象疾患名の記載や診断基準を満たすなどの明示的な情報のみであるか、それ以外のコンテキストを組み合わせているか、を調査する。たとえば腎細胞癌を対象疾患とした場合には、「腎細胞癌の患者」「腎細胞癌は認めない」といった記載から腎細胞癌の有無を医師が判断し

たならば、カルテ自然文における対象疾患名の明示的な情報のみを用いて判断したと考える。その一方で「腎臓に異常なし」の記載から腎細胞癌は無いと医師が推測した場合、対象疾患名が書かれていないため、この記載はコンテキストとする。(詳細は3章参照)

3. 手法

3.1 実験対象患者

2012年1月1日 - 12月31日の1年間に東京大学医学部附属病院(東大病院)外来受診あるいは入院した患者のうち、2009年 - 2014年の間に健診以外で2回以上受診した患者 104,522人。平均年齢 52.6(標準偏差 26.5)才、女性 57.2%。

3.2 対象とする疾患

対象とする疾患は下記の10疾患である。2型糖尿病、本態性高血圧、糖尿病性腎症、PBC(原発性胆汁性肝硬変¹)、クローン病、前立腺癌、乳癌、腎細胞癌、大動脈解離、肺塞栓。

病名は、Unified Medical Language System® (UMLS®) Metathesaurus® (2012AB Edition) から選択した。日本でエビデンスに基づいたガイドラインが存在する疾患を選択した²。Annotationなどの過程で、実際に患者を診察していない医師が、診断過程や治療内容の妥当性を評価するための根拠としてエビデンスに基づいたガイドラインを用いるためである。これらの条件を満たす疾患のうち、慢性疾患や急性疾患、診断や治療経過の判断に重要な検査値などの数値か画像所見か、患者数の多寡、の観点から、なるべく多様な疾患を対象とできるように対象疾患を選択した。

3.3 対象とする疾患毎の患者数の選定

目視でのannotationは時間がかかるため、EHRデータを確認する患者数を極力減らしながらも、全対象患者の中の正例の割合が母集団と同じであるようにサンプリングを行いたい。そこで、全対象患者の中での有病率が10%未満の疾患については、正例が全て入っていると考えられる集団を作成することで、annotation作業の対象患者数を絞った。具体的には、ガイドラインに登場した、検査項目または治療薬が一度でもオーダーされている症例、あるいは、病名やその略称、疾患に特徴的な症状のカルテ自由文への記載が認められる症例を選択した。これを仮母集団と呼ぶ。仮母集団の中からさらに、母集団の比率が予想される有限母集団に対するサンプルサイズの選定方法に基づいて計算された患者数をランダムに抽出し、サンプルとして採用し、annotationを行った[3]。信頼度係数(信頼率0.95に対応する正規分布点)は1.96、誤差の最大値は3%とした。そして、仮母集団のサンプルサイズに基づき全対象患者からランダムに抽出する患者数を決定した。サンプル数は疾患毎に異なり平均8193.0人(520人-30,710人)である。うちEHRデータを実際にannotationした患者数は平均461.9人(280人-650人)である。

¹ 2016年頃より世界的にPBCの正式名称は「原発性胆汁性肝障害」に変更されたが、本実験は2014年までのデータを対象とするため、旧来の名称を用いた。

² エビデンスレベルと推奨レベルが共に記載されており、診断および治療の双方について言及があるガイドラインを選択した。

3.4 実験手順の概要

ここで、行う実験手順とその意図の概要を説明する。本研究では2つの実験を行う。どちらの実験でも、カルテ自然文における対象疾患名の記載や診断基準を満たすなどの明示的な情報のみでe-phenotypingを行う。

最初の実験では、明示的な情報のみで疾患の有無を完全に分類可能な分類器の存在を仮定する。その基盤として、あらかじめ各患者について、対象疾患に関して先述の明示的な情報が記載されているか否かを、医師が目視で確認しておく(詳細な手順は3.4章)。その結果を用いれば、明示的な情報が存在する患者を正例、存在しない患者を負例と全例正しく分類できる分類器を仮定することができる。その分類器を用いたe-phenotypingで得られた精度は、すなわち、明示的な情報を抽出するという方針でe-phenotyping手法を開発する際に得られる精度の理論上の最大値である。仮に、あらゆる評価指標の理論上の最大値が100%の疾患ばかりであれば、e-phenotypingでは、用いるべきデータは疾患の有無に関係する明示的な情報だけで十分であり、より良い手法開発に注力すべきであると考察ができる。理論上の最大値が低い値の評価指標が存在した場合、その評価指標が高く確保できるe-phenotyping手法を開発するためには、様々なEHRデータを総合的に使うことが重要と考察できるだろう。このような実験の手順の詳細を3.5章で説明し、結果と考察は4.1章で示す。

次に、一部の疾患に対して、明示的な情報のみを用いたe-phenotyping手法を実際に作成し、得られる精度を計算する。このときに、上記の理論上の最大値を取れない場合には、その原因を考察する。作成した手法の不備が原因の可能性もあるだろう。しかしそれに加えて、利用可能なデータの不備が原因の可能性もあるだろう。たとえば、紙に書かれた他院紹介状のみに「2型糖尿病」と明示的に書かれているが、その他の経過記録には明示的な病名記載がない患者を考えた場合、この紹介状に書かれたテキストデータを電子的に利用できないと、e-phenotyping手法がこの患者を2型糖尿病ではないと判断してしまうだろう。この実験は3.6章で手順を詳細に説明し、結果と考察は4.2章で示す。

これら2つの実験の基盤として、対象疾患の有無を医師が目視で判断する際に用いる情報が、カルテ自然文における対象疾患名の記載や診断基準を満たすなどの明示的な情報のみであるか、それ以外のコンテキストを組み合わせているか、を、対象患者ごとにあらかじめ調査した。詳細は3.4章で述べる。

3.4 医師が目視で確認した内容

この研究では2名の医師が、各患者のEHRデータを目視で確認した。ここで用いたEHRデータは、東大病院の電子カルテシステムで閲覧可能な全データである。結果が異なる場合には2名で話し合い、結果を1つに定めた。確認した項目は、<1>カルテ自然文における対象疾患名の記載や診断基準を満たすなどの明示的な情報、<2>それ以外のコンテキスト、の2種類の情報の有無である。後述の通り、<1>は5カテゴリーの中のいずれが当てはまるか、<2>は3カテゴリーの中のいずれが当てはまるか、を各患者に対して確認する。

<1>については、<1-1>対象疾患名の記載がある、<1-2>対象疾患名の疑い記載がある、<1-3>対象疾患の検体検査結果で診断基準を満たす、<1-4>対象疾患についての明示的な情報がない、<1-5>対象疾患名に否定表現が修飾された

記載がある、の5カテゴリーのいずれが当てはまるかを判断した。2型糖尿病と本態性高血圧に関しては、それぞれ糖尿病または高血圧という上位疾患の半数以上を占める疾患であるため、上位疾患名の記載(疑い、否定、を含む)または診断基準を満たす際にも、上記の基準の対象疾患に当てはめて判断した。なお、この確認に際しては、医師に、記載の意味を自由に判断することが許されている。例えば、「クローン病」という記載があった場合に、患者がクローン病を持つという意味ではなく、鑑別疾患としての記載であると判断した際には、<1-1>として判断しない。

<2>は、<2-1>対象疾患(2型糖尿病と本態性高血圧に関しては、上位疾患も含める)の可能性を高める記載、<2-2>コンテキストはない、<2-3>対象疾患(2型糖尿病と本態性高血圧に関しては、上位疾患も含める)の可能性を低める記載、の3カテゴリーのうちいずれが当てはまるかを判断した。

たとえば2型糖尿病に関してある患者のEHRデータを確認した際、「糖尿病」と明示的な病名記載があれば<1>については<1-1>と判断し、その糖尿病がステロイド内服によるものだと判断されている記述が認められる場合には<2>について<2-3>と判断する。なお、各項目および確認手順の具体的な詳細は、先行研究[2]で筆者らが作成し公開しているannotation criteriaに基づいている。詳細は割愛するが、<1><2>の各カテゴリーの具体例を表1に示す。

表1 <1><2>の各カテゴリーの具体例。記載が存在しない<1-4>と<2-2>、検体検査の診断基準である<1-3>は除く。

<1-1>	「2型糖尿病」		
<1-2>	「2型糖尿病の可能性がある」		
<1-5>	「2型糖尿病ではない」		
<2-1>	「食事療法のみで治療中」		
<2-3>	「糖尿病はステロイドによるものだと考えられる」		

	<2-1>	<2-2>	<2-3>
<1-1>	(i)		(ii)
<1-2>			
<1-3>			
<1-4>	(iii)	(iv)	
<1-5>			

図1 医師が確認した情報に基づいた分類。<1>のカテゴリーと<2>のカテゴリーの組み合わせで、4種類に分割できる。たとえば、<1>については<1-4>、<2>については<2-1>が確認された患者は(iii)に分類される。

表2 医師が確認した情報に基づいた分類に基づいた、理論的な感度と陽性的中率の計算に必要な分類

	Gold standard	
	正例	負例
疾患の有無に関する明示的な情報のみで e-phenotyping を行うと仮定した際の結果	正例 (i)	負例 (ii)
	負例 (iii)	負例 (iv)

3.5 明示的な情報のみで e-phenotyping を行う際にとりうる精度の計算方法

この実験においては、コンテキストを用いる必要なく、明示的な情報のみで疾患の有無を完全に分類可能な分類器の存在を仮定した。これは図1で示す、<1-3>と<1-4>の境界での分類が精度100%で行える分類器を意味する。

その分類器を e-phenotyping 手法に採用して、疾患の有無の観点から gold standard を付与したデータに適用した際に取りうる感度、陽性的中率、特異度のそれぞれの最大値を計算する。計算方法について説明する。まず、疾患の有無に関する明示的な情報のみで e-phenotyping を行うと仮定した際の正例は(i)と(ii)、負例は(iii)と(iv)である(表2)。評価のための Gold Standard としては、図1の(i)と(iii)を正例、(ii)と(iv)を負例とした。Gold standard が医師の直観と反さないことは、EHR データを目視で確認した医師2名で確認した。たとえば(iii)に分類された患者は、明示的な情報のみで e-phenotyping を行うと負例に分類されるが、Gold standard としては正例である。このようにして、明示的な情報のみで作成した e-phenotyping 手法を用いてとりうる感度、陽性的中率、特異度の最大値は下記のようにまとめられる。

$$\begin{aligned} \text{感度} &= (i) \text{の患者数} / ((i) \text{の患者数} + (iii) \text{の患者数}) \\ \text{陽性的中率} &= (i) \text{の患者数} / ((i) \text{の患者数} + (ii) \text{の患者数}) \\ \text{特異度} &= (iv) \text{の患者数} / ((iii) \text{の患者数} + (iv) \text{の患者数}) \end{aligned}$$

この計算の意図をここで説明する。

仮に上記の e-phenotyping 手法と Gold standard の組み合わせで得られた感度、陽性的中率、特異度がすべて100%で変わらない疾患ばかりであれば、疾患の有無に関する明示的な記載を正しくテキストから抽出でき、また検体検査が診断基準を満たすかを正しく分類できる分類器を作成することが、疾患依存性の低い e-phenotyping 手法のフレームワークと言える。これは技術的には比較的難しい課題ではないだろう。

逆に、値が小さい評価指標があるとするならば、その評価指標を高値で確保できる e-phenotyping 手法を作成するためには、疾患の有無に関する明示的な情報だけをデータとして用いるだけでは不十分で、コンテキストなど多様な情報を用いるべきだと考えられる。

3.6 文字列抽出に基づく e-phenotyping 手法で用いたルール

前立腺癌、乳癌、腎細胞癌の正例の抽出について、文字列抽出を行う。この手法により、今まで述べてきた理想的な分類器がとる感度、陽性的中率、特異度の値(以後、理論的な値、と呼ぶ)を達成可能かどうか、実験を行った。本実験では、下記のルールに基づいて、文字列抽出を行った。この3疾患は、医師が検体検査結果を用いて疾患の有無を判断した症例、つまり明示的な情報として<1-3>を用いた症例がなかったため、本章で説明する文字列抽出で e-phenotyping を行うことと、疾患の有無に関する明示的な情報のみで e-phenotyping を行うことは同等である。

文の区切りは句点またはコロンを利用した。
 ”病名文字列”(表3)が存在する文中に”除外条件”(表4)を満たす文字列が存在しない患者を正例として抽出した。
 “病名文字列”と”除外条件”の定義は表3,4の通りである。

表3 ”病名文字列”を抽出するルール。3疾患に共通したルール(A)と疾患ごとに異なるルール(B)について、「(A)かつ(B)」を満たす。ただし、(A)と(B)が4文字以上離れている場合にはカウントしない。」の条件を満たす文字列を抽出した。

ルール(A)	(がん √ 癌 √ cancer √カルチノイド) U (◯良性 √ ◯benign √ ◯がんば)
ルール(B)	前立腺癌 前立腺 √ prostate 乳癌 (乳 √ breast) U (◯乳頭 √ ◯乳白 √ ◯乳腺外科) 腎細胞癌 (腎 √ kidney) U (◯腎内 √ ◯副腎)

表4 “除外条件” 下記のいずれか

否定記載	(A)または(B) (A)NO √ non √ ません √ ない √ なし √ なく √ 無い √ 無し √ 無く √ 否定 √ 考えにくい √ 陰性 (B)”病名文字列”と「-」または「一」の間隔が 3 文字以内の場合
家族歴記載	FH √ 家族 √ 父 √ 母 √ 両親 √ 祖父 √ 祖母 √ 子供 √ 息子 √ 娘 √ 子ども √ 甥 √ 姪 √ 妻 √ 夫 √ 姉 √ 妹 √ 兄 √ 弟 √ 奥さん √ 旦那さん √ 奥様 √ 旦那様 √ ご主人 √ 長女 √ 次女 √ 長男 √ 次男 √ 親戚 √ 親せき
疑い記載	要素 √ ? √ ? √ s/o √ 疑 √ 様 √ あるかも √ 心配 √ 可能性 √ リスク √ 示唆 √ questionable √ 境界型 √ check √ 検診 √ 健診 √ 除外 √ 除き √ 大丈夫 √ チェック √ 精査 √ 検査 √ 確認 √ 注意 √ 鑑別 √ 検討 √ 有無 √ 次回 √ 経過 √ 評価 √ ruleout √ r/o √ 一時的 √ 一過性
その他	systemWrote √ crtName √ timeStamp √ structureCR √ 初診_既往歴

4. 結果と考察

4.1 明示的な情報のみで e-phenotyping を行う際にとりうる精度の最大値

表5 ヒートマップを数値で表にする。明示的な情報のみを用いた際に取りうる感度、陽性的中率、特異度の最大値。

	感度	陽性的中率	特異度
2型糖尿病	97.2%	100.0%	99.6%
本態性高血圧	98.2%	90.9%	99.4%
糖尿病性腎症	100.0%	93.8%	100.0%
PBC	100.0%	81.3%	100.0%
クローン病	100.0%	87.5%	100.0%
前立腺癌	98.9%	82.3%	100.0%
乳癌	99.4%	97.8%	100.0%
腎細胞癌	100.0%	98.4%	100.0%
大動脈解離	100.0%	100.0%	100.0%
肺塞栓	100.0%	100.0%	100.0%

表6 e-phenotyping 結果

	感度	陽性的中率
前立腺癌	93.6%	87.1%
乳癌	92.1%	78.0%

腎細胞癌	84.1%	85.4%
------	-------	-------

表7 誤例の分類。■は利用可能なデータなどの改良が必要だと考えた項目。□は用いたルールの改善などで改良が可能と考えられた項目。◇はコンテキストを用いないと抽出できないために誤例に入った症例数

	誤負例	誤正例
前立腺癌	■ 電子的に利用できないデータに病名がある: 1例(pdf データ:1例) □ 英語表現などの処理の不備:3例 ◇ 1例	■ 字義以上の記載の除外漏れ:7例 □ 他疾患の除外漏れ:2例 □ 否定の除外漏れ:1例 □ 疑い記載の除外漏れ:2例 ◇ 1例
乳癌	■ 電子的に利用できないデータに病名がある: 3例 (pdf データ:3例、組織診レポート:1例) □ 否定語記載の処理の不備:7例 □ 疑い記載の処理の不備:1例 ◇ 1例	■ 字義以上の記載の除外漏れ:26例 ■ 家族歴、他人の話の除外漏れ:12例 □ 他疾患の除外漏れ:5例 □ 否定の除外漏れ:3例
腎細胞癌	□ 改行など前処理不足:3例 □ 否定語の処理の不備:1例 □ 英語表現などの処理の不備:5例	■ 字義以上の記載の除外漏れ:5例 ■ 家族歴、他人の話の除外漏れ:1例 □ 他疾患の除外漏れ:2例 □ 疑い記載の除外漏れ:1例

疾患の有無に関する明示的な情報のみで e-phenotyping を行うと仮定した際に取りうる感度、陽性的中率、特異度の最大値は、表5の通りである。明示的な情報のみで作成した手法を用いた際に取りうる最高の精度の最低値は、陽性的中率は81.3%(PBC)、感度が97.2%(2型糖尿病)、特異度が99.4%(本態性高血圧)となることがわかった。

感度がすべての疾患で100%とならなかった理由は、実際のEHRデータを確認すると、高血圧の病名自体は書かれていないが、コンテキスト(例:塩分を控えるようにという内容の食事指導の記載が認められ、ほかに塩分摂取を控える必要がある疾患の既往がない)から高血圧と判断する、といった症例が存在し(図1、表2の(iii)にあたる)、これらの症例を誤って分類したためである。しかし、その割合は低く、経過記録等の自然文から病名の文字列を正確に抽出し、検体検査の結果を正しくルールで再現できれば、97%以上の感度が確保できると言える。これは技術的には困難な課題ではないだろう。

疾患については、全患者を対象とした際には、Gold Standardの負例が多いため、特異度は多くの疾患で100.0%を達成できることがわかった。2型糖尿病および本態性高血圧で100.0%が取れなかった理由は、上記の通りの図1、表2の(iii)にあたる患者の割合が高かったためである。

その一方で、陽性的中率は、最低で感度や特異度と比較して十分な精度を確保できないことがわかった。これは、ある一時点の経過記録には「PBC」などと病名の肯定記載が確か

に存在するものの、他の診療記録等のコンテキスト(例:肝硬変の原因として PBC 以外の原因の可能性が高いことに言及されている、主治医も判断に迷っていることが読み取れる、患者の状態の変化に応じて主治医の判断が変化している、など)から総合的に PBC ではないと判断した症例などの存在が原因である。このような症例は図 1、表 2 の(ii)にあたる。これらの症例は、対象疾患についてある一文における事実性判定が必要となるだけではなく、仮にそれが正しくできたとしても、他の鑑別疾患や類似する病態を持つ疾患の事実性判定まで必要となり、さらにそれらの疾患が対象疾患と類似していることや併発しえるかといった外部知識の参照や、それらのコンテキストの総合的な含意も判断する必要がある。これらは現状の自然言語処理技術では、個別に解決されていないタスクを複数組み合わせる必要がある課題であり、解決は難しいと考えている。そのため、EHR データを幅広く用いて機械学習などを行うことが有用だと考えている。実際に、疾患の知識を問わずに EHR データを可能な限り用いた機械学習で高い精度を確保したという報告も存在する[5]。このような研究の方向性が、疾患特異的なものではなく、一般的な方向性として妥当であることを確認できた。

また、最初に文字列抽出などで比較的簡単に十分な感度で対象患者を絞り込み、その後、絞った患者に対して機械学習などを適用して高い陽性的中率で患者を抽出する、といった 2 段階の手順を踏むことが、正例の抽出に有用とも考えている。機械学習は理論上、対象患者の正例と負例の比率が近いほど高い精度が得られることが知られており、疾患という、全人口を対象としたときには罹患率が高くて 1 割前後と考えられる対象を抽出する場合には、対象患者を絞り込む手順が有用だろうと予測される。

4.2 文字列抽出に基づく e-phenotyping 手法で得られた精度

前立腺癌、乳癌、腎細胞癌の正例の抽出について、文字列抽出により、感度および陽性的中率を再現可能か、予備的な実験を行った。文字列抽出はルールで記載した。結果は表 6 の通りで、表 5 に記載された、理論的な値は達成できていない。誤分類の例は表 7 の通りである。

まず感度が理論的な値を達成できなかったこと、つまり誤負例について検討する。現在、東大病院にて電子的に扱うことが容易なデータは経過記録であったが、それ以上のデータの記載を扱えないことの影響は感度の値に直して 0%(腎細胞癌)-1.7%(乳癌)程度しか影響しなかった。EHR データの情報の二次利用がさらに可能になる環境を整えなくても、技術的な改良が十分に行えれば現時点で理論的な感度を確保できることを示唆している。

陽性的中率が理論的な値を達成できなかったこと、つまり誤正例について検討する。字義以上の記載の除外漏れ(例:「クローン病」と否定表現の修飾なく記載されているが、それは患者がクローン病であることを意味しておらず、鑑別疾患などを意味している)、および家族歴や他者の罹患について、これを正しく判断できれば 6.9%(前立腺癌)-18.2%(乳癌)の値で陽性的中率が改善する可能性がある。これらを正しく判断させるための研究開発には、このような記載に対しての事実性判定のカルテコーパスの作成[6]が必要だと考える。

また前立腺癌については、鑑別疾患や家族歴の記載など、本来なら抽出されるべきでない記載が抽出され、明示的なデータだけを用いた e-phenotyping 技術では抽出されないはず

の症例が正例として抽出されたために、理論的な陽性的中率よりも高い値が得られるという結果になった。これは用いた技術の改善などにより改良可能であると考えられる。これは、精度の数値だけで評価を行うと、技術的な誤りを見落としかねないという教訓的な例である。

5. 本研究の限界と汎用性、将来性

本実験の問題点として、3.4 章で用いた、医師が確認する EHR データ内の情報の分類基準を、各疾患の専門医などの意見を仰がずに決定したことがあげられる。調べた限りで直接応用できるような先行研究が存在しなかったため、本実験に参画した医師 2 名の経験に基づいた分類となっている。そのため今後は、医師やその他医療者が、実際に臨床現場を想定した際に、どのような EHR データに基づいて患者の対象疾患の有無を判断するのか、認知科学的な実験などを行うべきだろう。

また、EHR データの記入者に依存した結果が生じている可能性も否定はしきれない。本実験では、記入者によって EHR データを実験対象から除外する(例:医学部生の実習における記録を除外する)ことは行っていないため、結果の汎用性は高いと考えている。しかし、診療科によって EHR データ記入についての指導や慣習が異なるなどの可能性は検討できていない。また他施設で同様の実験を行った際にも、医療体制の違いなどから結果が異なる可能性は否定できない。

本研究結果は、既存の e-phenotyping 手法からは独立して、しかも疾患依存性が低い e-phenotyping 開発フレームワーク作成のための基盤的な結果として応用できると考えている。また、研究より先に利用可能なデータを広げることが優先するのか、その必要はなく技術開発に取り組むのか、についても判断材料を与えると考えている。

6. 結論

e-phenotyping の観点から EHR データの特徴を整理し、それに基づいて、計算機処理における課題や用いるべき技術の選択、整備が必要なデータの有無を検討した。

端的な病名記載を抽出する、または、対象疾患の診断基準を満たす、が感度も特異度も 100%で正しく判断できる分類器を仮定し、それを実データに適用すると、全疾患で 97%以上の感度、99%以上の特異度が理想的には確保できることがわかった。このときに、電子的に扱うことが容易ではないデータの記載を扱えないことの影響は感度の値に直して最大で 1.7%(乳癌)程度であることを確認した。このことから、テキストから対象とする病名を正しく抽出する自然言語処理技術、および診断基準に当てはまる検体検査結果を抽出するための厳密なルール記載さえ確立すれば、感度や特異度は十分高い値を確保可能であること、経過記録が電子的に利用可能であれば、詳細な部門データや紙データをスキャンしたデータまで電子的な二次利用を至急進める必要性は低いことが示唆された。

しかし陽性的中率については、最低で 81.3%(PBC)と、上記の技術だけでは十分な精度を確保できないことがわかった。病名が明示的に記載されていたとしても、医師はその前後のコンテキストを用いて、その記載が実際に疾患の有無を意味しているか否かを判断している症例が多いことが原因であった。医師が判断しているような様々なコンテキストを計算機処理によって抽出することは、現状の自然言語処理技術では困難なタスクである。そのため、高い陽性的中率を確保するためには、テキストや診断基準に用いる検体検査データに限ら

ずに EHR データを幅広く総合的に用いた手法を開発すべきだと考えられる。

7. 倫理的配慮・謝辞

東京大学大学院医学系研究科・医学部倫理委員会により承認を得ている(承認番号 10791)。本研究は MEXT 科研費 16J05555 の助成を受けた。

参考文献

- 1) G. Hripcsak and D. J. Albers, Next-generation phenotyping of electronic health records, *JAMIA*, 20, 117-121, 2013
- 2) Rina Kagawa, Yoshimasa Kawazoe, Emiko Shinohara, Takeshi Imai, Kazuhiko Ohe, The impact of “possible patients” on phenotyping algorithms: Electronic phenotype algorithms can only be reproduced by sharing detailed annotation criteria., *Studies in Health Technology and Informatics* (in press)
- 3) 東京大学教養学部統計学教室編. 統計学入門. 東京大学出版会, 2008: 229-230
- 4) M. Huan et al., Desiderata for computable representations of electronic health records-driven phenotype algorithms, *JAMIA*, 22, 1220-1230, 2015
- 5) R.J. Carrol et al., Portability of an algorithm to identify rheumatoid arthritis in electronic health records, *JAMIA*, 19, e162-e169, 2012
- 6) 矢野憲, 若宮翔子, 荒牧英治, 医療テキスト解析のための事実性判定と融合した病名表現認識器, 言語処理学会 第 23 回年次大会 発表論文集, 2017