

医師の暗黙知に基づくタグ付きカルテコーパス作成の要件の検討

香川 璃奈¹ 篠原 恵美子² 河添 悦昌² 今井 健¹ 大江 和彦¹

¹ 東京大学大学院医学系研究科 ² 東京大学医学部附属病院

{kagawa-r, emiko}@m.u-tokyo.ac.jp, {kawazoe, ken, kohe}@hcc.h.u-tokyo.ac.jp

1 背景と関連研究

医療現場において、医療過誤の6割が診断の誤りとされる [1]。過度な細分化へ反省から医療者に従来以上に幅広い知識が求められる今日、医療行為の責任を負う立場にある医療者が納得できる、医療者の思考過程に基づいた診断支援システムが望まれる。医療機関では医師法などに基づき、診療に関する事項が必ずカルテに記録される。カルテには血液検査などの検査結果やCTなどの画像に加え、診断された病名、検査結果の解釈や、診断や治療方針の決定に至った理由がテキストで記録される。そこで世界的にカルテの電子化と研究利用が進む今日 [2]、診断に至る思考過程に関連する情報を付与したカルテコーパスの作成と利用が診療支援に有用だと考える。

我々は先行研究において、患者が持つ疾患の有無のアノテーション基準を作成した [3]。これは、病名を含むテキストだけでなく、病名を含まないテキスト、テキスト以外の検査結果や処方箋の構造化データも等しく確認する基準である。この基準を用いた場合と用いない場合で得られた正例数が異なるだけではなく、正例集合の臨床的な性質に有意差が認められた。このことから、病名以外のテキストや構造化データも用いた、医師の知見に基づいた詳細なアノテーションの重要性が示唆された。そこで、診断に利用する情報の付与にあたり、どのようなタグを使うべきか、何を対象データとすべきか、医師によるカルテデータの意味の質的な分類に基づいて検討する。

関連研究の検討にあたり、ある患者が肺塞栓症(肺動脈に血栓が詰まる疾患、エコノミークラス症候群として知られる)の診断に至るまでとその後を示したカルテ(図1)を例示する。実在データに基づくが結果に影響のない範囲でデータの詳細を変更した。図1では肺塞栓症の診断という観点から医師の直感に基づいて、肺塞栓症の診断に関連する14個のデータに分割した。既存の疑似カルテに対するコーパス作成のガイドライン [4]には病名や症状に付与するタグが規定され

ており、図1に青太文字で示した病名にモダリティを付与できる。図1の全てを総合的に判断してこの患者には肺塞栓症の罹患歴がある、とアノテーションした我々の先行研究 [3]と比較して、2010年に肺塞栓症に罹患したが2013年には罹患していないとわかる利点がある。しかし肺塞栓症の診断に関連するデータには、病名を含まないテキスト ((5), (9), (11), (12), (14))、検体検査結果 ((8)のd-dimer) や処方 ((4), (7)のヘパリン)の構造化データも含まれ、これらのデータの持つ情報を付与する必要がある。そのため、病名へのタグ付けだけでは目的を達成できない。カルテコーパス作成以外の分野も参照したが、本研究で目的とするのはテキストに限らないデータと作業者の理解との含意関係を付与する作業であり、表層や言い換え [5]からは得ることができない。また、同じCT画像の結果に基づき(6)(10)では肺塞栓症とは判断できないとしているが、(3)では肺塞栓症だと判断している。(3)を書いた医師は肺塞栓症の専門科である呼吸器内科所属である。このような、同一の検査結果や所見に対する解釈が医療者の専門性によって異なることは対話における言外の情報の扱いにも類似する。しかし、データの受け手毎に異なる解釈を総合した最終的な判断1つの付与が必要な点、すでに下された診断を後ろ向きに理解する作業なので作業者が医師であれば高い妥当性が期待できるという点で、雑談対話 [6]とは異なる。このような異なる解釈を扱う上で、カルテデータを閲覧した医療者が疾患の有無を事実と考える度合の”幅”を扱うべきだと予測される。しかし既存の確実性のアノテーション基準 [7][8]は、確実性を表す表現に基づいて分類するため構造化データを扱えない。

2 手法と実験対象

医師(医師免許取得6年目)が電子カルテデータを目視で確認した。臨床現場で医療者が用いる電子カルテの画面を用いて、データを全て閲覧し患者像を把握し、対象疾患の有無の判断の観点から図1のようにカルテデータを直感的に分割し意味を分類した。対象患

者を実際に診察しカルテ入力を行った医療者への意見聴取は行わなかった。3つの実験を行った。

病名以外の、病名を想起させる意味の分類：病名以外に疾患の有無を想起させるデータの意味を分類した。

病名の表記と意味の差異の分類：病名について、表記通りのモダリティを持たない場合の意味を分類した。病名は、疑いあるいは否定を意味する表現(品詞を問わない)が修飾しているかを基準に、肯定記載、疑い記載、否定記載の3種類に分類した。

病名変化が存在する症例数：医師による解釈の違いや診断に至る思考過程が記録されているカルテには、病名の変化が存在すると予測される。たとえば、患者が初めて病院を受診したときには疑い記載が記録され、その後検査を行い診断が確定すると疑い記載が書かれなくなり肯定記載が書かれる。同一の患者のカルテに肯定、疑い、否定記載の少なくとも2種類以上が登場する、すなわち病名記載が変化する症例が占める割合を確認した。

対象疾患は2型糖尿病、本態性高血圧、糖尿病性腎症、原発性胆汁性肝硬変、クローン病、前立腺癌、乳癌、腎細胞癌、大動脈解離、肺塞栓症の10疾患である。慢性疾患や急性疾患、診断に重要なのは検査値などの数値かテキストで表現する所見か、患者数の多寡、の観点から多様な疾患を選択した。実験対象患者は、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%)から疾患ごとにサンプル数の患者をランダムに抽出し(平均461.9人(280人-650人))実験を行った。1症例あたりカルテテキストは平均110ファイル、行われた検査データは1,060データ。

3 結果

病名以外の、病名を想起させる意味の分類

病名を含まないテキストや、テキスト以外の情報について、疾患の有無を推測できる情報の意味は11種類に分類でき(表1)、これらのデータが存在した患者は全ての対象疾患で認められた。可能性を高める情報と低める情報を含む患者が占める割合は10疾患平均でそれぞれ13.5%と15.3%と、表1の情報を付与しないと、有無を問わず疾患の想起を誤りうるということがわかった。表1は、カルテデータから想起できる対象疾患の確実性の度合という観点に基づき、上の行ほど対象疾患の可能性が高まるデータ、下の行ほど可能性が低まるデータと解釈する。たとえば図1(14)は、肺塞栓症に特異的な検査結果の否定を意味する。さらに、同一

日のデータまたは同一文の中に肺塞栓の有無について一見矛盾するような情報も存在する。たとえば(6)は肺塞栓症の否定記載だが、同一日の(7)では血栓を治す薬剤のヘパリンを処方している。(5)(11)が肺塞栓症を完全には否定しきれない確実性である上に肺塞栓症は致命的な疾患であるため、肺塞栓症であると仮定して治療を開始したと解釈できる。否定記載(6)の存在にもかかわらず(5)(11)から肺塞栓症を完全には否定しきれないという確実性の解釈には、肺塞栓の病態や、血栓が生じるとd-dimerの検査結果が異常値を示しうる、という医学知識が必要である。これらの結果から、分類を付与すべき情報は文法的な基準では一概に決定できず、医学知識に基づいて分割すべきだと考えられた。一方で、画像データに情報を付与する症例はいずれの疾患でも認められなかった。これは、画像で認められた所見が画像レポート、あるいは経過記録などのテキストの形で記載されるためだと推測された。

ここで図1の(12)を説明する。この記載はDVTという主として下肢の静脈に血栓ができる疾患を実質的に否定する情報だが、これは表記だけでは理解できない。仮に交通事故の患者を診察して骨折を疑うときに同じ記載が認められたならば、骨折がないことを示唆する情報と解釈することも可能であろう。DVTがあると下肢の痛みが生じうる、また、DVTの血栓が肺動脈まで移動すると肺塞栓症が生じるという医学知識に加えて、手術などではDVTと肺塞栓症を併記してリスク管理することが多いため、肺塞栓症に言及している(9)-(11)の次に記載されている(12)はDVTを意味するであろう、といった臨床現場における経験知を用いると、この日に肺塞栓症に罹患していないことをより確実に判断するための補助的な役割を果たしうる。このことから、医学知識のみならず臨床現場における経験知が有用なことが示唆され、分類の付与が外部知識参照だけで十分に行えないことが示唆された。

構造化データとテキストの双方を用いないと、情報の意味を正しく判断できない症例が全ての対象疾患で認められた。図1(5)(8)はその1例であり、d-dimerが基準値を超えていることの原因が何か、すなわちどの疾患について考えるときにd-dimerの検査値を用いるべきなのかは(5)にテキストの形で記載されている。また、図1(3)(4)(5)は記録日が異なるデータであるが、これを互いに参照すると、(7)で投与を開始したヘパリンが肺塞栓症に対しての処方であることをより確実に判断できる。このことから、複数のデータを自由に組み合わせることで分類を付与する必要性が示唆された。さらに、ある1つのデータに対する分類の付与の

ために、前後のなるべく全てのカルテデータを示すことでより正確な作業につながる事が示唆された。

病名の表記と意味の差異の分類

肯定、疑い、否定記載の3種類に分類した病名が含意する情報は表記通りの意味および家族歴、表記の間違い [4] 以外にも認められた。たとえば、図1の(1)の肯定記載は既往歴(過去の罹患)を意味する。(13)の肯定記載は、(10)で肺塞栓症が否定されており、ともに列挙されている「肥満」「無気肺」「SAS」という疾患の臨床的特徴から、低酸素血症(図1(9))の原因の鑑別疾患としての記載であり、この時点での肺塞栓症の罹患を意味しないと判断できる。鑑別疾患とは、診断が確定する前に可能性のある疾患として列挙する疾患である。鑑別疾患を必要十分に列挙することは医師の基本技能である。さらに鑑別疾患として列挙されたあとに言及がない場合には、その疾患が実質的に否定されたことが推測できる。つまり、(13)で肯定記載として列挙された4疾患すべてに鑑別疾患というモダリティを付与することは、各疾患の有無の確実性の度合の判断に必要である。これが正しく判断可能なことは、別の医師が診察を引き継いだ際に、過去の医療行為を正しく理解して追加で行う必要十分な検査や治療方針を決定することに有用である。しかし(13)で列挙された疾患が鑑別疾患であることは表記には現れず、前後から想起される患者像と、対象疾患以外の関連する疾患の医学知識を組み合わせて判断する意味であり、既存の事実性や確実性のアノテーションの枠組み [8][9] だけでは意味の付与は行えないと推測される。また図1には示していないが、疑い記載は診断の観点から考えると、“確定的な診断を下す前に疑われる疾患”だけでなく、“患者が高齢で診断を確定するために必要な検査が行えず、あえて診断を確定しないで治療を継続している場合の疑い記載”などが存在することが確認された。このように、肯定、疑い、否定記載の持つモダリティがさらに複数種類に分割できることが推測され、病名が持つ複数種類の意味の分類が診断に利用する情報として必要であると考えられた(投稿準備中)。

病名変化が存在する症例数

検査結果や専門医による解釈による病名変化が認められた割合は10疾患の平均で2.6%(0.5%-7.8%)であり、病名を想起させる病名以外の情報を含む患者の割合と比較すると少数であった。このことから、下された診断を後ろ向きに理解するためには病名以外の情報の利用が必要で、つまりそのデータに情報を付与する必要性が確認された。

4 考察

4.1 カルテコーパス作成の要件

本実験から、前後の日付の全てのカルテデータをデータ種別を問わず確認して患者像を正しく想起し、病名以外のデータが持つ意味や、病名が持つモダリティを理解する作業には対象疾患以外の病名や検査や治療の医学知識、さらには臨床現場の経験知の必要性が明らかになった。日本語で整備された標準病名マスター V4.03[10]の病名基本テーブルには25,676病名が存在する。本研究では対象外であった、症候群と言われる多数の症状や所見をまとめた疾患概念も対象とすると、包含される症状や所見の分類にも多様性が生じうる。このことから作業者は医師、可能な限り専門医が望ましいと考える。そのためクラウドソーシングは利用が難しく、報酬も高く設計する必要性が推測される。その一方で、タグを付与するデータに文法的なルールを策定しなくても、作業者の解釈に基づいたデータ分割を期待できる利点があげられる。

上記を踏まえて、実現可能と考えられたことを整理する。まずは対象患者を絞り、その患者のカルテデータ全てに分類を付与する。用いるタグは表1に基づく。画像に情報を付与する症例が見つからなかったことから、構造化データは全てテキストに書き下す。カルテ内に分散するデータをまとめて情報を付与する場合は、同時参照すべきデータであることを示すタグを付与する。タグを付与するデータは、名詞句、文、文章など文法的なルールを策定せず、作業者が自由に解釈した意味に基づいて分割する。我々は表1の分類、および病名が表記と異なる意味を持つ症例が占める割合を疾患ごとに確認しており(投稿準備中)、予備的に作成した基準に基づく実験を行うことが今後の課題である。相互に参照すべきデータについては、データを多く示しすぎることによる作業の質が下がる可能性が考えられ、前後の日数どの程度のデータを示すべきか、今後実験を行いたい。

4.2 実験の限界

医師1名で分類を行ったことは本実験の限界である。しかし、専門医や各カルテデータの記録者の意見をあおぐと、さらに詳細データを分類できると予測される。大規模な教育病院である東大病院のデータのみを用いた解析であることも限界点である。しかし、より記載が曖昧になりやすいと予想される他病院の電子カルテデータを用いた場合、本研究で指摘した問題点はより大きくなると予測される。すなわちこれらの限界によって本研究の意義は変わらないと考えている。分

類の網羅性や排他性は保証できていないものの、対象疾患のうち6疾患は分類者を含む医師4名が病名と表1の分類を用いて各患者の疾患の有無のアノテーションを行い、他の情報を用いる必要がある症例が存在しないことを確認し、医師間のアノテーションの一致度は κ 係数0.70-0.98であった。分類の定量的な評価のために、consensus based methodの利用や認知科学的な実験[11]の実施も検討中である。しかしカルテデータの理解は臨床現場で身につける暗黙的な知見であり、本実験で明らかにしたような具体的な分類は臨床現場では明示的に利用されないため困難が予想される。予備的な基準に基づく実施と並行して実験を行うことが現実的であろう。

5 おわりに

医療者の思考過程に基づいた診療支援システムに利用するタグ付きカルテコーパスの作成を目的として、電子カルテデータを医師が目視で確認し、10疾患の診断に関連するカルテデータを分割し、データの意味を分類した。コーパス作成では、本研究で明らかにしたタグを構造化データとテキストを包括的に扱い意味的に分割したデータ、さらに互いに参照した複数のデータに付与すべきだと明らかになった。病名も文法的な意味にとどまらない分類を付与すべきだと推測された。データの付与には、医学知識と臨床現場での経験知が対象疾患を問わず必要になる。これは既存の評価用データの作成枠組みで対応できないと考察され、マルチモーダルなデータに経験的な専門知に基づく分類を付与する方法論の発展が待たれる。

謝辞・倫理的配慮

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)。JSPS 科研費16J05555の助成を受けた。

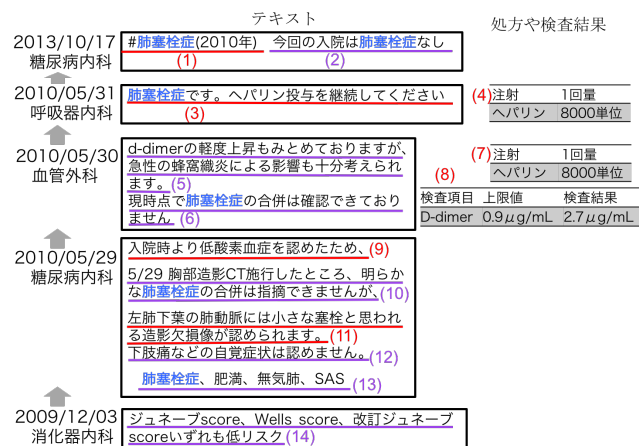


図1: 肺塞栓症の診断に関するカルテの例。データ番号は筆者注、肺塞栓症の罹患が推測されるデータ(赤)と罹患の否定が推測されるデータ(紫)に大別できる

参考文献

- [1] GD Schiff et al., NEJM, 1066-1069, 2010
- [2] WQ Wei et al., Genome Medicine, 7, 41-54, 2015
- [3] R Kagawa et al., Studies in health technology and informatics, 245, 432-436, 2017
- [4] NTCIR11 MedNLP2 Annotation Guidelines ver1.0, <http://mednlp.jp/ntcir11/guideline.pdf> (2018年1月13日閲覧)
- [5] 金子ら, 自然言語処理学会第19回年次大会発表論文集, 928-931, 2013
- [6] K Mitsuda et al., IWSDS2017, 1-9, 2017
- [7] Medlock B and Briscoe T, ACL2007, 992-999, 2007
- [8] 川添ら, IPSJ SIG Technical Report, 77-84, 2009
- [9] NTCIR13 MedWeb コーパス ガイドライン ver2.0, (2018年1月13日閲覧)
- [10] 標準病名マスター, <http://www.dis.h.u-tokyo.ac.jp/byomei/index.html> (2018年1月13日閲覧)
- [11] R Iida et al., LAW7&ID, 214-222, 2013

表1: 病名以外のデータが持つ意味

大分類	中分類	小分類
対象疾患の可能性を高める	対象疾患に間接的に言及	対象疾患に特異的な治療内容、検査結果、症状
		鑑別疾患を否定する検査結果
対象疾患の可能性を低める	対象疾患を間接的に否定	対象疾患と同じ薬剤が処方される他の疾患が存在
		鑑別疾患が存在する可能性が高い記載、検査結果
		対象疾患の部分的な否定
	対象疾患を実質的に否定	対象疾患が生じる臓器に他の疾患が存在
		対象疾患が生じる臓器に異常がない記載や検査結果
		可能性があると言及されているのに、検査結果に言及なく治療を行っていない
		将来発症の可能性
上位概念にあたる疾患の否定		
対象疾患に特異的な症状や検査結果の否定		