

# 医師の経験をよく反映した擬似カルテ文書の作成方法

香川璃奈 馬場雪乃 鶴嶋英夫  
筑波大学医学医療系 筑波大学システム情報系

{kagawa-r, hideo-tsurushima} @ md.tsukuba.ac.jp baba@cs.tsukuba.ac.jp

## 1 背景と目的

電子カルテには、日々の臨床現場での多様なデータが蓄積される。カルテの中でも文書には、患者の主観的な訴えや医師の思考過程など、構造化が容易とはいえない多彩な情報を、医師が日々記載しており、その再利用は教育研究目的でも有用である。しかしその一方で、文書に記録される豊富な情報は病歴すなわちプライバシーであり、保護せねばならない。名前や住所などの個人識別符号を除去した文書であっても、制限のない公開は日本を含む多くの国と地域で難しく、ドイツなどでは研究者による利用が事実上不可能となっている [1]。

そこで我々は、時代や地域に依存する個人情報の定義、言語や医療体制、対象疾患を問わずに本物らしい擬似的なカルテ文書データセットを作成できる、汎用的な枠組みが必要だと考えた。

### 1.1 提案フレームワークの方針

まず、教育研究目的で有用な疑似カルテ文書データセットが満たすべき要件を2つに整理した。

#### (1) 臨床的妥当性

カルテ文書には、患者の訴え、検査結果、治療計画とその根拠、家族歴や嗜好歴などが記載される。疑似文書には、これらが1人の患者ごとに、臨床的な矛盾なく記載されている必要がある。また、実臨床で臨床家に対応に苦戦する、複雑な患者像を描けてこそ、教育研究目的で価値がある本物らしいデータセットになると考える。その一方で、臨床的妥当性は医療制度などにも影響されるため、実際にどのような臨床的特徴を持つ患者がどの程度存在するのかが不明である。

#### (2) カルテ文書に特徴的な記載の妥当性

多忙な臨床現場において専門知識を持つ医師が記載するカルテ文書には、文脈によって意味が変化する極端な省略表現やスペルミスを含む非文法的な構造が多用されることが、国や地域を問わず指摘されている[2, 3]。その一方で、どのような特徴が存在す

るのかを網羅的に明らかにした研究は存在しない。

これらは、カルテ文書の網羅的な閲覧が倫理的観点から不可能であることから、医師にとっても経験的な知見であり評価することは難しい。Generative Adversarial Network に代表されるデータ生成技術は医療データにも応用されているが、現状で生成されている文書は限られた内容の短文[4, 5, 6]が主であり、これらを組み合わせても2つの要件を満たす疑似文書データセットの作成は困難である。そもそもプライバシー保護の観点から、データ生成に資する大規模なカルテ文書収集も容易ではない。また、既存の公開カルテ文書として、医師が記憶と経験に基づいて記載した670個の文書[7]や、医師が教育目的で記載した書籍からの抜粋[1]から成るデータセットが存在するが、これらは先述の2つの要件の評価が明示的に行われておらず、特に(2)の特徴が反映されていないことが指摘されている[8]。すなわち2つの要件を満たして自由に利用可能な疑似カルテ文書は存在しない。

そこで我々は2つの要件を満たすデータセット作成のために3つのステップからなるフレームワークを提案する。

Step 1: 臨床的妥当性を満たす文書の大規模な生成

Step 2: カルテ文書に固有な特徴の種類とその多様性の顕在化

Step 3: Step 1 で生成された文書のうち Step 2 で得られた特徴を満たす文書を選択し、疑似カルテ文書データセットとする。

### 1.2 本研究の目的と貢献

本研究は、提案フレームワークの有効性を実証することを目的とする。効率的かつ現実的な手法で実証するため、クラウドソーシングを利用する。

本研究の貢献は以下の通りである。

- (i) 専門的な文書を生成するフレームワークの提案
- (ii) カルテ文書に固有な特徴の種類とその多様性の顕在化
- (iii) 医師が擬似的に記載する文書よりも本物のカル

テ文書に類似していると評価された、9,856 個の擬似カルテ文書の公開

## 2 方法

### 2.1 材料

筑波大学附属病院を 2013 年 1 月~2018 年 9 月に受診した患者 53,246 人からランダムに選択した 245 人の患者(女性 48.2%、平均年齢 60.1 才)の文書(カルテ文書の一般的なテンプレートである SOAP フォーマットに則るもの。以下、本物のカルテ文書と呼ぶ)を利用した。また、既存の擬似カルテ文書[7]から SOAP フォーマットに則る 257 個の文書(以下、既存公開文書と呼ぶ)を利用した。各文書が記載された診療科を筆頭著者が推定した。

解析には以下を利用した。Python 3.7.2、NumPy 1.17.2、SciPy 1.2.1、Pandas 0.24.2、Matplotlib 3.0.3、NetworkX 2.3、Apyori 1.1.1、Community 1.0.0、R 4.0.2、exactRankTests packages 0.8-21。

### 2.2 提案フレームワークの実現例

本研究では、クラウドソーシングを利用することで提案フレームワークを低コストで実現した(図 1)。前処理として、本物のカルテ文書の固有名詞や日付などを筑波大学附属病院臨床研究倫理審査委員会から承認を得た手順で書き換え、患者個人が特定されないことを著者ではない 3 名が確認した。

#### step 1 : 臨床的妥当性を満たす文書の生成

模倣をすることで背景知識がなくても事象を再現できる人間の認知特性[9]を参照し、クラウドワーカーが本物のカルテ文書を模倣することで臨床的妥当性を満たす文書を生成した。(a) まず、本物のカルテ文書を SOAP フォーマットに則って S, O, A/P の 3 種類に分割した 260 個の文書(140 人の患者分)を作成した。(b) その中から同じ診療科で記載された同じ意味(S または O または A/P)に分割された文書をワーカーに提示し、ワーカーは提示された文書を模倣した文書を 1 つ記載した。一度に提示される文書の数は 1, 2, 3 個からランダムに決定した。ワーカー 1 人あたり 20 回作業を行い報酬は 660 円とした。(c) ワーカーが記載した模倣文書のうち模倣の際に提示された本物のカルテ文書に同じ患者が含まれる模倣文書を SOAP フォーマットに則って再構成した。これを生成文書と呼ぶ。同じ患者に基づく模倣文書を再構成することで、生成文書の中で臨床的な矛盾が

生じにくい一方で、別の患者のデータに基づく記載が含まれることで非実在患者の情報が記載された文書であることを担保した。

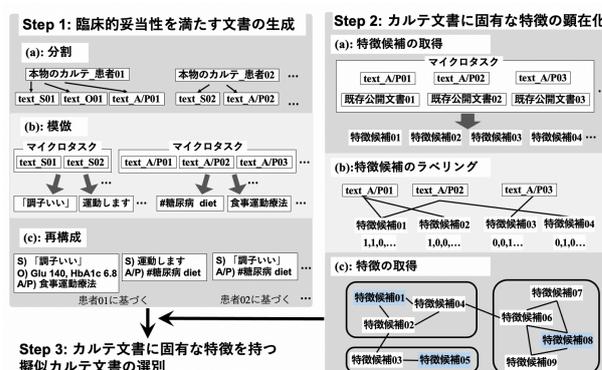


図 1 提案フレームワークの概略

#### step 2 : カルテ文書に固有な特徴の顕在化

比較を行うことで相違点の判断能力が高まるといふ人間の認知特性[10]を参照し、以下の 2 段階のマイクロタスクを実施した。(a) 本物のカルテ文書と既存公開文書をそれぞれワーカーに提示し、2 者の違いを「患者の不安な気持ちが記載されている」「日本語の文章の中に英単語が混ざっている」のように任意の数の yes/no クエスチョンのかたちでワーカーが記載した。一度に提示する文書数は 1, 3, 5, 7, 10 個ずつのいずれかからランダムに決定した。得られた回答を特徴候補と呼ぶ。72 人のワーカーが 50 回ずつ作業を行い、1 人あたり報酬は 550 円とした。(b) 類似する特徴候補を集約することで、カルテ文書に固有な特徴を明らかにしたい。そこでまず、ランダムに選択した 100 個の本物のカルテ文書と、その文書が満たす特徴候補のペアを取得した。本物のカルテ文書一つと特徴候補がワーカーに提示され、ワーカーは提示された本物のカルテ文書が満たす特徴候補を任意の数、選択した。1 人あたり 25 回作業を行い報酬は 440 円とした。得られた本物のカルテ文書と特徴候補のペアのうち、出現頻度が低くても強い相関があるペアを取得するため、lift[11]が 5 以上のペアを採用した。(c) 類似する特徴候補を集約するため、(b) でペアとして採用された本物のカルテ文書集合間の Jaccard 係数が 0.5 以上の特徴候補同士を類似しているとみなした。特徴候補をノードとし類似する特徴候補間にエッジがあるネットワークを作り、Louvain 法を適用してコミュニティを抽出した。コミュニティ毎に、構成する特徴候補の中から延べ 105 人のワーカーの多数決で決定された最も代表的な特徴候補を、本物のカル

テ文書に固有な特徴(以下、記載特徴と呼ぶ)とした。1人あたりランダムに選択された50個のコミュニティについて作業し報酬は440円とした。各記載特徴について、カルテ文書を業務で閲覧する治験コーディネーター3名が「この特徴を持つカルテ文書は存在しない(-2)」から「存在する(+2)」までの5段階のリッカート尺度で評価した。

### step 3 : カルテ文書に固有な特徴を持つ擬似カルテ文書の選別

生成文書から100個をランダム選択し、生成文書一つと記載特徴を延べ50人ワーカーに提示した。ワーカーは、提示された文書が満たす記載特徴を任意の数、選択した。1人あたりランダムに選択された10個の生成文書について作業し報酬は550円とした。記載特徴のうち治験コーディネーターによる評価の平均が0より小さい記載特徴を満たすと判断したワーカーが半数以下の生成文書を、2つの要件を満たした擬似的なカルテ文書と判断した。

## 2.3 作成した擬似カルテ文書の評価

提案フレームワークによって得られた疑似カルテ文書と、ランダムに選択した100個の本物のカルテ文書との類似性を評価した。比較対象として、2人の医師が臨床現場を再現した実験環境で電子カルテ様画面を利用して記載した19個の文書(医師記載文書と呼ぶ)および257個の既存公開文書と本物のカルテ文書との類似性の比較も行った。類似性の量的評価として、文書あたりの、標準病名数[12]、McCab-0.996とmecab-ipadic-2.7.0 dictionaryを利用して求めた形態素数、文章の難易度(readability)を評価する13段階からなる指標であるT-13[13, 14]を計測した。類似性の質的評価として、医師3人が、上記4種類の文書からそれぞれランダムに選択した12個ずつをランダムな順番に閲覧し、各文書を実際のカルテ文書だと思うかを「思わない(1)」から「思う(10)」までの10段階リッカート尺度で回答した。各指標について、本物のカルテ文書とそれ以外の3種類の文書の間、スコアの平均値の有意差検定を行った。Shapiro-Wilk検定とBartlettの結果に基づき、文書あたりの標準病名と形態素数の平均値の比較にはBrunner-Munzel検定を用い、readabilityと医師による質的評価のスコアの平均値の比較にはMann-Whitney U検定を利用した。また、データセットのスコアの分布間のKullback-Leibler (KL)擬距離(式(1))を求めた。 $p(x)$ は本物のカルテ文書のスコアの分

布、 $q(x)$ はそれ以外の3種類のカルテ文書のスコアの分布である。

$$D_{KL}(p(x) \parallel q(x)) = \sum_x p(x) \ln \frac{p(x)}{q(x)} \quad (1)$$

## 3 結果

### 3.1 クラウドワーカーの統計データ

延べ1,662人のクラウドワーカーのうち97.3%は医療従事者でなく、医療系の学生経験もなかった。

### 3.2 提案フレームワークの実現例

#### step 1 : 臨床的妥当性を持つ文書の生成

35人のワーカーが延べ700個の文書を作成し、再構成により9,856個の文書が生成された。

#### step 2 : カルテ文書に固有な特徴の顕在化

3,197個の特徴候補から176個の記載特徴が抽出された。コミュニティが抽出され、3人の治験コーディネーターの評価によりそのうちの166個がカルテ文書の記載特徴とされた(表1)。また、Step2と同様の実験を3人の治験コーディネーターが行い取得された105個の特徴と、記載特徴を著者が比較したところ定性的な違いは認められなかった。

#### step 3 : カルテ文書に固有な特徴を持つ擬似カルテ文書の選別

100個の生成文書のうち83個が、2つの要件を満たした擬似的なカルテ文書と判断された。

表1 カルテ文書に固有な特徴の例

日本語の文章の中に英単語が混ざっている
他科の医師への相談内容の記載がされている
検査数値が数字・英語のみで記されている
患者の苦しみが発言通りに書かれている

### 3.3 作成した擬似カルテ文書の評価

既存公開文書、医師記載文書、擬似カルテ文書、本物のカルテ文書のそれぞれについて、文書あたりの標準病名数(平均値)は23.56、37.47、45.18、60.48、形態素数(平均値)は89.54、105.47、164.04、178.34、readabilityスコア(平均値)は9.70、8.08、10.22、10.29、医師の質的評価(平均値)は4.53、5.25、5.20、5.0であった。既存公開文書や医師記載文書と異なり、擬似カルテ文書では4つの全ての指標について、スコアの平均値が本物のカルテ文書と等しいという帰無仮説が棄却されなかった(図2, 表2)。各指標の本物の

カルテ文書のスコアの分布からの KL 擬距離は、既存公開文書、医師記載文書、擬似カルテ文書の順に、文書あたりの標準病名数は 0.385、0.569、0.244、形態素数は 0.312、0.942、0.115、readability スコアは 0.098、4.160、0.633、医師の質的評価は 1.573、1.111、0.767 であった。すなわち、擬似カルテ文書が他の 2 つのデータセットよりも本物のカルテ文書に類似していると判断した。作成した擬似カルテ文書は公開<sup>i</sup>した。

表 2 作成した擬似カルテ文書の評価

	95% CI	P value	95% CI	P value
	標準病名数		形態素数	
擬似カルテ文書	0.39 to 0.56	0.54	0.45 to 0.62	0.36
医師記載文書	0.25 to 0.56	0.20	0.18 to 0.50	0.046
既存公開文書	0.22 to 0.34	4.2e-11	0.25 to 0.38	1.3e-08
	Readability		医師の質的評価	
擬似カルテ文書	-6.3e-05 to 1.0+e0	0.54	-2 to 2	0.36
医師記載文書	1.2e-05 to 4.0+e0	0.02	-2 to 1	0.046
既存公開文書	5.2e-05 to 1.0+e0	0.03	-1 to 2	1.3e-08

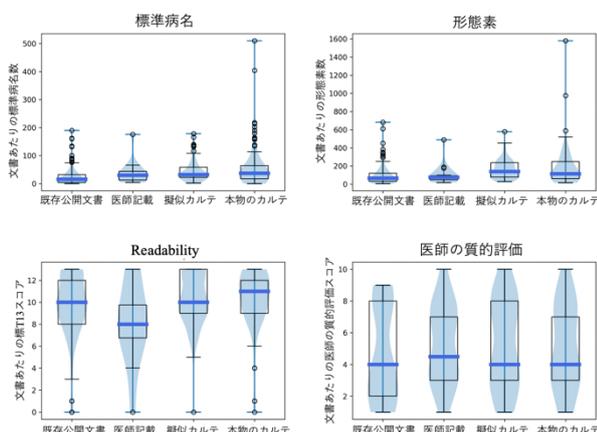


図 2 各評価指標のスコア

## 4 考察

本研究では、医師にとっても経験的なカルテ文書の特徴に着目して提案した擬似文書作成フレームワークを利用して、医師が電子カルテ画面を実験用に

再現して記載した文書よりも本物のカルテ文書に類似している擬似カルテ文書を作成した。

### 4.1 提案フレームワークの汎用性と限界

提案フレームワークは、国や地域に依存しない点で汎用性があり、本実験のようにクラウドソーシングを利用すれば少数の本物のカルテ文書から擬似的なカルテ文書を作成できる利点を持つ。その一方で、カルテ文書に固有の特徴を明らかにする際に、実際のカルテ文書には存在しない特徴を明らかにすることはできない提案手法である点は限界である。

### 4.2 擬似カルテ文書の汎用性と限界

本研究で作成した擬似カルテ文書は、プライバシーの問題なく自由に利用可能である。日本語の擬似カルテ文書として、公開文書数は知る限り最大である。文書には「ビタミン B12 内服=痺れ」(痺れの訴えに対してビタミン B12 を内服する、という意味)のようなカルテ文書に特徴的な記載が存在した。

一方で、医療の背景知識を持たないクラウドワーカーによる記載、また利用した本物のカルテ文書の特徴を原因とする未知の偏りが生じる可能性は否定できない。カルテ文書の網羅的な閲覧が倫理的観点から不可能であることに起因する限界である。

## 5 結論

本研究では、文書の内容の一貫性を保つ必要があり、文書に固有の特徴が明らかになっておらず、大規模なデータ収集が困難、という 3 つの特徴を持ったカルテ文書を、擬似的に生成するフレームワークを提案した。さらに提案フレームワークを、専門知識を持たないクラウドワーカーにより実現することで、医師が臨床現場を模した環境で記載したカルテ文書よりも本物らしいと判断できる擬似カルテ文書を作成した。カルテ文書をはじめとした公開が困難な文書資源を利用した言語処理研究において、提案フレームワークが利用できれば幸いである。

## 6 倫理的配慮および謝辞

本研究は筑波大学附属病院臨床研究倫理審査委員会の承認を得た(H30-145)。JST 未来社会創造事業(JP19211284)、JSPS 科研費(JP19K19347)、筑波大学社会工学データコモンズの助成を受けた。

<sup>i</sup> <https://zenodo.org/record/4064153>

## 参考文献

- [1] Lohr, Christina, Sven Buechel, and Udo Hahn. "Sharing Copies of Synthetic Clinical Corpora without Physical Distribution—A Case Study to Get Around IPRs and Privacy Constraints Featuring the German JSYNCC Corpus." *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. 2018.1259-1266.
- [2] Savkov, A., Carroll, J., Koeling, R. et al. Annotating patient clinical records with syntactic chunks and named entities: the Harvey Corpus. *Lang Resources & Evaluation* **50**, 523–548 (2016).
- [3] Aramaki A, Imai T, Miyo K, Ohe K. Robust classification for ungrammatical and fragmented texts. *Proceedings of the 23<sup>rd</sup> IEEE International Conference on Data Engineering Workshops*; 2007 Apr 15-20:195-201
- [4] Lee SH. Natural language generation for electronic health records. *NPJ Digit Med* 2018; 1(1): 1-7. PMID: 26900011 doi: 10.1016/j.artmed.2016.01.003
- [5] Guan J, Li R, Yu S, Zhang, X. Generation of synthetic electronic medical record text. *Proceedings of the 2018 IEEE International Conference on Bioinformatics and Biomedicine*; 2018 Dec 3-6:374-380; Madrid, Spain. Washington, DC: IEEE; 2018. doi: 10.1109/BIBM.2018.8621223
- [6] Ive J, Viani N, Kam J, Yin L, Verma S, Puntis S, et al. Generation and evaluation of artificial mental health records for natural language processing. *NPJ Digit Med* 2020; 3(1): 1-9. PMID: 32435697 doi: 10.1038/s41746-020-0267-x
- [7] Aramaki E, Morita M, Kano Y, Ohkuma T. Overview of the NTCIR-11 MedNLP-2 task. *Proceedings of the 11th NTCIR Conference*; 2014 Dec 9-12:147-154; Tokyo, Japan. Tokyo: NII; 2014.
- [8] 香川璃奈, 篠原恵美子, 河添悦昌, 今井健, 大江和彦: 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、言語処理学会第 24 回年次大会論文集 2018. 757-760
- [9] Kuniyoshi Y. The science of imitation — towards physically and socially grounded intelligence —. *Real World Comput Proj Joint Symp* 1994; 95-96.
- [10] Gentner D, Markman AB. Structure mapping in analogy and similarity. *Am Psychol* 1997; 52(1): 45-56. doi: 10.1037/0003-066X.52.1.45
- [11] Tan PN, Kumar V, Srivastava J. Selecting the right objective measure for association analysis. *Inform Syst.* 2004; 29: 293-313. doi: 10.1016/S0306-4379(03)00072-3
- [12] ICD10 対応標準病名マスター(引用日: 2020 年 9 月 15 日) <https://www2.medis.or.jp/stdcd/byomei/index.html>.
- [13] Sato S, Matsuyoshi S, Kondoh Y. Automatic assessment of Japanese text readability based on a textbook corpus. *Proceedings of the 6<sup>th</sup> International Conference on Language Resources Evaluation*; 2008 May 29-30:654-660; Marrakech, Morocco. Luxembourg: ELRA; 2008.
- [14] 帯 3 日本語テキストの難易度を測る(引用日: 2020 年 9 月 15 日) <http://kotoba.nuee.nagoya-u.ac.jp/sc/obi3/>

## 付録 公開した擬似カルテ文書の例

(S)

プールで週1回のリハビリ、10分の散歩を初めて、ストレスは減り、体が軽くなったようだ。体調はまあまあだ。

(O)

推定塩分摂取量 6.11g/day

体重 66.7kg

骨格筋 23.2kg

骨格筋量 25.4kg

低血糖の自覚はないです。

74-199-177-

(A/P)

昨夜からランタス3単位で開始したところ、今朝から低血糖になり、ランタスは1単位へと減らす。

また、退院は血糖コントロールが安定するまで外科主科で診てもらおう。

当科と併せて診てもらい、血糖コントロール続けていく。

退院の目処が立ち次第外科へ連絡。

Q(1-1-1)G(0-0-0-1↓)へ

ストレスがあり、栄養や運動の調整は難しい

ビタミンB12内服＝痺れ