

[別添 3]

平成 30 年度厚生労働科学研究費補助金 政策科学総合研究事業

(臨床研究等 ICT 基盤構築・人工知能実装研究事業 総括研究報告書)

カルテ情報の自動構造化システムと疾患数理モデルの逐次的構築, 及び, 自動構造化機能を有した
入力機構の開発

研究代表者 荒牧英治 奈良先端科学技術大学院大学 研究推進機構

研究要旨

電子カルテは患者情報が全て記録されているものの, 非文法的かつ断片化した表現が多く自然言語処理を応用した利活用は困難であった. これを二次利用するため申請者等 (申請者荒牧及び分担者河添が所属する研究室主宰者の大江ら) は, 2008年から電子カルテから医療用語の自動抽出及び自動コーディングを行う研究に従事してきた. その成果は, 日本内科学会の症例報告検索システムなどとして実用化され, 現在も用いられている. 本研究は, 電子カルテの二次利用のさらなる実用化に向けて問題となる次の2つの課題を解決する.

(課題1) 実用化可能な解析精度の達成

(課題2) 電子カルテに組み込み可能な実装の開発

若宮翔子 (奈良先端科学技術大学院大学 研究推進機構・特任助教)

河添悦昌 (東京大学大学院医学系研究科・特任准教授)

A. 研究目的

これまで, 医療医学用語をまとめる試みは数々なされてきた. これらは主に医師・研究者などの医療者が使う用語を対象としており, 電子カルテなど医療現場で扱う技術は向上しつつある. 万病辞書もその辞書の一つであり, かつてない多くの用語を収載している. その一方で, 近年増えつつある患者が記述したテキストを扱うためには十分な用語が収載されていない.

本研究はこれまで予定より早く医療用語のデータの収集が進んでおり, 今後は, コストの許す限り人手による精査を進めていくのみが課題であると考えている. そこで最終年度である本研究は, 当初の計画にはなかったこの患者表現をも一部収載しようと発展的な研究を行った.

患者によるテキストは, 「闘病記」, 「病の語り」, 「当事者ブログ」, 「患者 SNS」, 「patient narrative」など, さまざまな呼び方が存在するが, 本書では, 「患者テキスト」と呼ぶことにする. この患者テキストに記述される内容は, 不安などの心の問題 (36%), 症状・副作用・後遺症 (16%) 家族, 周囲の人との関係 (10%), 告知やインフォームドコンセント・セカンドオピニオン (9%), 診断・治療 (8%) であり [医療言語処理 (荒牧英治著) より], 不安と並んで, 医学的な情報も多い.

患者テキストを利用した自然言語処理研究は始まったばかりであり, その課題は明確に定まっていない. しかし, 最初に解くべき問題として, 医学知識が豊かであるとは限らない患者が記述する患者表現を正式な医学表現に変換することが挙げられる. これは, 従来, 自然言語処理で表記ゆれ解消と言われた技術と近いが, 単なる語の言い換えレベルにとどまらず, 語から句への変換が必要となる場合もある.

例えば, 医療用語としては<感覚鈍麻>として記述されるべき症状が, 患者テキストにおいては「指先がピリピリする」と記述されうる場合がある. このように, 擬音語や擬態語 (以降,

擬音語と擬態語を総じてオノマトペと呼称する)を用いた動詞句レベルや文レベルでの表現が頻用される。患者テキストを医療現場で活用するためには、これらの非医療用語を言い換える必要がある。

このような言い換え全般をカバーすることは現在の自然言語処理では困難であり、大規模な辞書が必要であると考えている。このような背景の中、本研究は、大規模な患者表現の辞書を構築する。

B. 研究方法

B-1. 症状オノマトペの収集

前述したように患者表現が医学表現と異なる点は次の2つである。

- ・語彙的ギャップ：擬音語擬態語の頻用、特にオノマトペに代表される表現。
- ・構造的ギャップ：複合名詞である医学用語が動詞句として表現される。

これら2つの違いを考慮して、オノマトペに特化した収集方法と、クラウドソーシングを用いて動詞句を含んだ表現を募集する方法の2つを併用して表現を収集し、標準病名との対応をとる。以降、前者を症状オノマトペ、後者を症状句と呼称する。

B-2. 部位リストから部位+オノマトペ表現の自動収集

Google n-gramコーパスから「(部位)が(オノマトペ)する」という表現を収集する。

擬音語、擬態語はカタカナであるかどうかで判定する。人体部位は事前に作成した表1を用いた。

この結果、部位、オノマトペ、頻度の3つ組のリストが得られた。次に、これを精査し、不適切なものを除いたデータを構築した。収集したデータのイメージを把握しやすくするため、人手で可視化したものを図1に示す。

B-3. オノマトペ表現の標準化

次にオノマトペ表現がどのような医学表現に対応するかを人手により紐づけた。

医学表現としては、標準病名とし、部位+オノマトペのペアで標準病名との対応を得た。例えば、「頭・ガンガン」は標準病名<頭痛>、「耳・ガンガン」は<耳鳴>となる。曖昧性があり一意に決定できない場合は、複数の標準病名を列挙した。

表1：人体部位リスト（抜粋）

人体部位	オノマトペ
頭部	がんがん
大腿部	パンパン
上腕	プルプル
膝	バキバキ

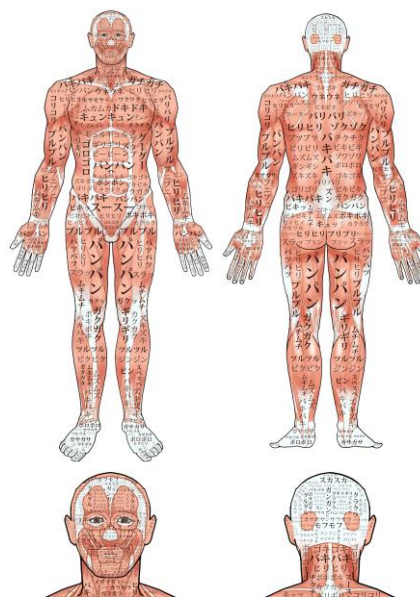


図1：部位ごとに可視化したオノマトペ

B-4. 症状句の収集とその方法

オノマトペと異なり、句の表現の形式は「AのB」「AがBする」「AとBがCする」など様々であり、事前に形式を決める収集はできない。患者表現を医学表現に紐付けるのではなく、逆に、医学表現を患者表現（形式は問わない）に言い換えることで収集を行った。医学表現としては<標準病名>を用い、クラウドソーシングにて収集した。収集方法としては、クラウドソーシングでは万病辞書の表現形から、内科学会頻度が25以上のものを対象に100名にアンケートを実施した。その際の教示は以下の通りである。

「症状を表す各単語について、他にどのような言い方（表現）ができるかを書いていただくタスクです。特定の地域や集団のみで使用されている表現（方言や隠語など）でもかまいません。またいくつ書いていただいてもかまいません。ただし、本調査は医療用語に関するものですので、それ以外の分野で同様の表記があるものに関しても、医療用語として認識してください。」

同一表現のものは一部省略（例：’すりガラス陰影’と’スリガラス陰影’）し、411単語のアンケートを9部に分け、性別、年齢、出身地と、図2のような設問ごとに自由記述をしてもらった。収集したデータは句読点を削除し、同一表現ごとに集計したものを降順に並べた。漢字、ひらがな、カタカナは全て別表現として捉えている。参加者の性別、年代の内訳は図3に示すのとおりである。



図2 自由記述フォーマット

性別	人数
男性	568
女性	332
年代	人数
10代	6
20代	82
30代	195
40代	353
50代	193
60代	58
70代	12
80代以上	1

図3 参加者の性別と年代

（倫理面への配慮）

本研究については課題名「電子的診療録の自動構造化機能を有した入力ソフトウェアの開発研究」で、奈良先端科学大学院大学情報学系の倫理審査に申請し、申請が受理されている。

C. 研究結果

症状オノマトペ・症状句の収集の結果を、ウェブ上 (<http://sociocom.jp/~data/2019-pde/>) に掲載した。また、本収集方法を論文として投稿予定である。

D. 考察

評価の困難さ：

本研究では、オノマトペに注目した方法とクラウドソーシングを用いた方法という2つの方法により、患者症状表現の収集を行った。患者表現収集の大規模な試みの事例は研究代表者らの知る限りなく、また、既存のリソースもないため、結果の評価は困難である。

例えば、クラウドソーシングで症状の表現を収集して本リソースの網羅性を評価しようとしても、これはクラウドソーシングによる手法と同じことであり、同じデータを再現してしまうだけである。今後は、構築方法とともに評価方法を検討する必要がある。

応用可能性：

本研究は患者症状表現の収集を行った。患者表現収集の大規模な試みの事例は乏しく、どれくらいをカバーすれば全体の何%をカバーするのかわからず、また、何が実現できるのかわからない。評価方法とともに小規模な応用を繰り返しながら、検討することも有効であると考えている。潜在的な応用先は次のようなサービスを考えている：

(1) スマートフォンやスマートスピーカーを用いた患者症状の抽出。

患者が日常的に用いるデバイスに日々蓄積される自然文から、患者の症状の抽出を行い、想定外の有害事象やアンメットニーズの発見につなげる。

(2) 待ち時間の問診票

病院の待ち時間などに患者に問診票を記載する際に、その自然文を解析し、カルテに転送するなど、病院業務の軽減につなげる。

(3) 医療者-患者間コミュニケーション支援

患者表現と医療用語を結びつけた辞書により、患者と医療者の双方が、相手側の用語を知り、コ

コミュニケーションを円滑にする教育効果が期待できる。

E. 結論

本研究では、これまで大規模な収集が困難であった患者の症状表現の収集を行った。これは現在、ウェブ (<http://sociocom.jp/~data/2019-pde/>) にて公開している。

F. 健康危険情報

特になし。

G. 研究発表

1. 論文発表

- 該当なし

2. 学会発表

- Kaoru Ito, Hiroyuki Nagai, Taro Okahisa, Shoko Wakamiya, Tomohide Iwao, Eiji Aramaki: J-MeDic: A Japanese Disease Name Dictionary based on Real Clinical Usage, LREC 2018. (Miyazaki, Japan)

H. 知的財産権の出願・登録情報

特になし