

令和 6 年 6 月 4 日現在

機関番号：14603

研究種目：基盤研究(B)（一般）

研究期間：2021～2023

課題番号：21H03543

研究課題名（和文）患者テキストを臨床に活かす医療言語処理基盤の構築

研究課題名（英文）Medical Language Processing Platform for Clinical Use of Patient Texts

研究代表者

荒牧 英治（Aramaki, Eiji）

奈良先端科学技術大学院大学・先端科学技術研究科・教授

研究者番号：70401073

交付決定額（研究期間全体）：（直接経費） 12,000,000円

研究成果の概要（和文）：患者と医師を繋ぐ自然言語処理基盤を構築する。本研究では、乳がん患者の副作用把握や肝疾患患者の禁酒指導といった具体的な課題で、患者テキストの表記ゆれ吸収や事実性判定を実現する医療言語処理基盤を構築し、情報処理の医学的応用の新展開を模索する。昨年度は、患者のテキスト記録を外来受信時に医療者へ送信する自然言語処理基盤を備えたアプリを開発した。今年度は、禁酒におけるコントロールにて、アプリを用いて、医療者の負担を増やさずに医療の質を向上させることができるかを実証する実験を行い、論文発表した。

研究成果の学術的意義や社会的意義

患者アプリは多くあるが、医師が診療時に用いるアプリケーションにおいて、自然言語処理が利用されているものは少ない。これは、臨床現場の高い要求に、エラーも含まれる自然言語処理をどう用いるかの良いデザインが足りていないからかと思われる。本研究は参画している医師（関西医科大学消化器内科、京都大学乳腺外科）の協力のもと、臨床現場での利用を重視して、研究をデザインしている。今後、自然言語処理を始めとした情報技術の臨床応用のロールモデルになることを目指している。

研究成果の概要（英文）：We will build a natural language processing infrastructure that connects patients and doctors. In this study, we will construct a medical language processing platform that can absorb distortions in patient text and determine the factuality of the text for specific issues such as understanding the side effects of breast cancer patients and providing guidance on abstinence from alcohol to patients with liver disease, and explore new developments in the medical application of information processing. Last year, we developed an application with a natural language processing platform that transmits a patient's text record to a medical care provider at the time of outpatient reception. This year, we conducted an experiment to demonstrate whether the application can improve the quality of medical care without increasing the burden on the medical staff in the control of alcohol abstinence, and published a paper.

研究分野：自然言語処理

キーワード：自然言語処理 医療情報学

1. 研究開始当初の背景

我々は医療 AI の大きなボトルネックが医療テキストの解析 (以降、医療言語処理と呼ぶ) だと考えている。近年、大規模データを用いた機械学習によるアプローチが急速に医療を変えつつある。画像診断 AI はすでに上市され

表 1: 医療 AI が抱えるジレンマ。

	優先したい内容	制約事項	患者から得られるデータ量
AIの要請	患者情報が多いことが望ましい		増加が必須
医療者の要請	治療効果を上げたい	ただし、医療者の負担を増加させたくない	現状を維持
患者の要請	できるだけ診てほしい		増加を希望
コロナ禍の社会要請	できるだけ来院回数を減らしたい	ただし、医療の質は下げない	減少の恐れ

つあり、AI ベース創薬加速やロボット技術を活かす手術支援など様々な領域で高度な情報技術が導入されている。しかし、かかりつけの医院にて、患者と医師のコミュニケーションを助けてくれるような日常の診療現場で活躍する AI は未だに存在しない。診療場面の AI サポートは、医師の偏在や高齢者の増加といった医療リソースに貧窮する日本にとって大きなニーズがあると見込まれる。ではなぜ、診療場面の AI 進出は進まないのでしょうか。一つの原因は、医療現場の多忙さがある。現場の実感としては、将来的に AI によって便利になると分かっているが、今、AI 開発にさける時間はないというのが本音であろう。つまり、大量のデータを用意すれば開発が可能な AI (将来の便益) と、現状の負担を増やせない医療側 (現在必要なコスト) との間にジレンマがある (表 1)。さらに、COVID-19 以降の社会では、患者が病院に来る回数を減らしたいが、医療の質は下げられないという社会からの要請も加わり、強制的にこのジレンマの解消が迫られている。

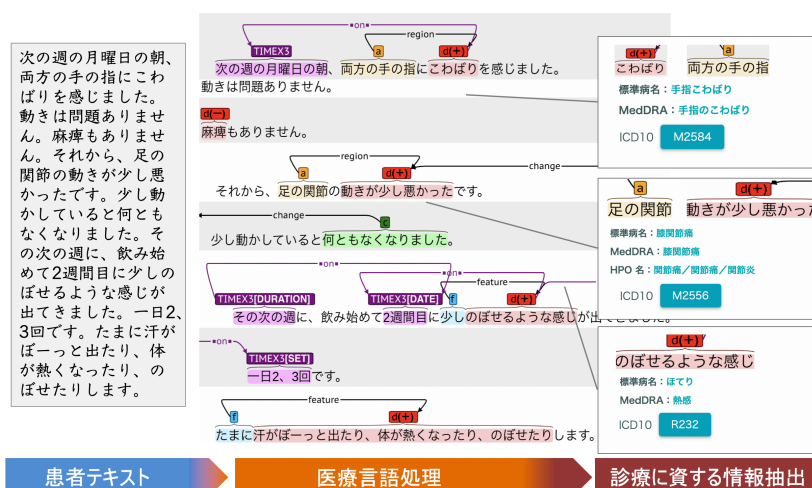


図 1: 本研究が実現する患者と医師を繋ぐ自然言語処理基盤。患者テキストから診療に必要な情報を自動抽出する要素技術一式 (固有表現認識, 事実性判定, 表記ゆれ吸収, 特定のイベントを情報抽出) を研究開発する。

我々は、このジレンマを解く突破口となるのは、患者情報を要約して医療者へ繋ぐAIだと考える。本研究で開発する基盤は、まずは患者から膨大な情報を可能な限り集め、次に、そこから医療者が必要な情報のみをフィルターし表示し、判断の一助とする。この基盤は、法整備やインフラ整備を必要とせず、医療者側からはゼロコストで診療に使える形態である。

ただし、このAIを実現するためには、時として非文法的かつ断片化した患者テキストを解析し、症状の変化や有害事象イベントといった臨床上重要な情報を高い精度で抽出する必要がある。これには、単なる用語抽出（固有表現認識）だけでなく、用語が示すイベントが実際に発生したのかどうか（事実性判定）、さらに同義語、表記ゆれの吸収など経て特定のイベントを情報抽出するといった一連の自然言語処理の総合タスクを解く必要がある。図1はこれを端的に示す例である。自由記載した日記に含まれる副作用を日時（頻度）、部位、症状の国際標準分類（ICDやMedDRA）に変換し、データベースに格納する。このような技術は、医師が記述する電子カルテにおいてはすでに確立しつつあり、読影所見など領域を絞れば精度は85%（F値）を超えている[AIP-創薬 PRISM 報告書]。その一方で、この技術を患者テキストに適用した例は少なく、体系だった研究報告は国際的にも少ない。本研究は、本邦で初めて本格的な患者の記述テキストに対しての医療言語処理技術を確認し、臨床での実証を行う。研究の枠組みを図2に示す。この枠組は、患者側からみれば、AIと病院が一丸となり自分を治療してくれていることになり、医療者側からは、患者がAIにより自助している形となる。患者の膨大な情報を入手しつつも、来院回数を増加させず、治療効果を高める可能性があり、コロナ後の社会の要請にもマッチする。

2. 研究の目的

患者情報を要約して医療者へ繋ぐことができる自然言語処理基盤を構築し（自然言語処理研究としての目的）、医療者の負担を増やさずに医療の質を向上させることができるかを実証する（臨床研究としての目的）。この検証には、2つの臨床研究（抗がん剤の副作用と禁酒におけるコントロール）を用いる。

1. 医療言語処理基盤の構築：患者テキストに対して、時間表現、部位表現、症状表現、精神的症状（不安や悩み）を自動抽出し、国際標準のターミノロジーに紐付けるシステムを開発する。
2. 臨床研究による検証：30人規模での臨床研究を行う。患者から抽出した情報を医師端末へ送信し、結果を参照しつつ診療を行う。評価は、治療が奏功したかどうかを調査する。実験デザインは、観察研究の形をとる。乳がんシナリオにおいては、既存の紙ベースの副作用調査表をベースにした治療法との比較を行う。肝疾患シナリオにおいては、再飲酒率をア

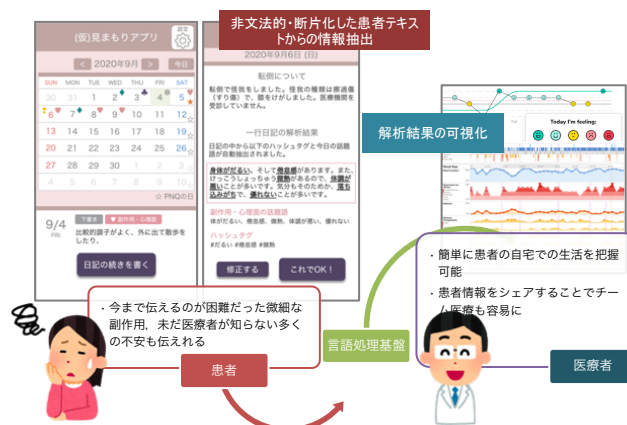


図2: 提案する枠組み。患者側アプリ（左）と病院側システム（右）を提供する。患者は日々の記録を記述し、医師はそのサマリーを閲覧する。副作用の詳細な把握など医師が把握でき、精密な医療に結びつく。患者にとっても、日々の記録を正確に担当医に伝達することができ、安心・安全に繋がる。

ウトカムとして、従来治療と比較する。

3. 研究の方法

【研究項目1：医療言語処理基盤の構築】

患者テキストに対して、時間表現、部位表現、症状表現、精神的症状（不安や悩み）を特定する。時間表現、部位表現、症状表現については、系列ラベリング技術によって抽出を行い、表記ゆれ吸収の処置を経て、国際標準のターミノロジーに紐付ける。精神的症状については、文単位で既存の分類体系（静岡分類）をベースにしたラベルを付与する。奈良先端科学技術大学院大学はシステム開発を担当し、臨床研究用のシステムだけでなく、一般の患者が利用できるシステムを公開する。

【研究項目2：乳がんでの実証実験】

乳がん治療中の患者の治療において、動的に変化する副作用は、時として命に関わる。副作用の特定には、多様な症状、例えば「口内炎が多い」といった患者報告が重要な手がかりとなり、自由記述から副作用情報を抽出することは合理的である。これまでは、紙面でのアンケート（末梢神経障害質問表）が用いられてきたが、本研究では、質問表では捉えきれない症状を患者の記述から検出することを目指す。

【研究項目3：肝疾患での実証実験】

アルコール依存症の患者については、禁酒は必須であるものの、患者本人のみでこれを継続することは困難である。継続には、頻回な介入が有効であるとされているが、頻回なサポートは医療リソースを逼迫し、現実的には十分なサポートは困難である。本研究では、低医療リソースなサポートを実現するため、患者に禁酒生活の詳細（飲酒量、何が原因で飲酒したか、またその時の精神状態）を自然言語で記述してもらい、そこから自動抽出した情報を参考に診療を行い、きめ細かい患者コントロールを実現する。

4. 研究成果

当初の目標は、臨床研究用のシステムだけでなく、一般の患者が利用できるシステムを公開することであった。目標の半分である臨床研究用のシステムについては、研究項目1で達成できたと考えている。残り半分の患者用のシステムについても研究項目2と項目3において、実施している。このように、全ての研究項目について以下のように一定の成果があった。

【研究項目1：医療言語処理基盤の構築】

患者テキストに対して、時間表現、部位表現、症状表現、精神的症状（不安や悩み）を特定するシステムに関しては、開発を終了し、MedTXT-CR-JAとして公開を行った。これを用いた応用として、医薬品を対象としたCRESTバイオDX領域を開始することができた。このCRESTプロジェクトにて、MedNERNというプログラムとして現在も開発を続行している。さらに、このMedNERNを現在、様々な企業共同研究やプロジェクトに応用しつつある。このように、目標の一部である臨床研究用のシステムについては、研究項目1で達成した。

【研究項目2：乳がんでの実証実験】

乳がん治療中の患者の治療にアプリの配布を行った。これはePRO（イープロ；electronic Patient Reported Outcome）と呼ばれる従来から存在するアプリと近いが、本研究の新規チャレンジとして、自然文についての記述も備えている。これを、京大において進行中の末梢神経障害の観察

研究にアプリを用いて 100 名以上のデータが現在も集まりつつある。ただし、自然文の記述は少なく、患者記述を促すことが大きな課題であることも判明した。

【研究項目 3 : 肝疾患での実証実験】

アルコール依存症の患者に、京大同様にアプリを導入した。1 点異なる点は、医師が外来時にアプリの結果を用いて診療を行う点である。アプリを導入した結果、一定の効果を得ることができ、論文発表[1]を行った。サンプルサイズは小さいものの、自然言語処理研究を臨床に導入する事例は過去多くなく、今後の研究の発展のために必要な重要な成果だと考えている。特に、患者の脱落の理由であるアプリ操作の難易度、臨床での使用の簡便さなど多くのノウハウが得られた。

以上のように 3 つの研究項目にて、全ての共同研究組織が本研究に参画し、一定の成果を上げたと考えている。協力していただいた医師、患者の皆様に深く感謝している。

参考文献

[1] Noriyo Yamashiki, Kyoko Kawabata, Miki Murata, Shunichiro Ikeda, Takako Fujimaki, Kanehiko Suwa, Toshihito Seki, Eiji Aramaki, Makoto Naganuma: Narrowing the Patient-Physician Gap Based on Self-Reporting and Monthly Hepatologist Feedback for Patients With Alcohol-Related Liver Disease: Interventional Pilot Study Using a Journaling Smartphone App, JMIR formative research, 7:e44762, 2023 (2023/12/19)

5. 主な発表論文等

〔雑誌論文〕 計2件（うち査読付論文 2件/うち国際共著 0件/うちオープンアクセス 2件）

1. 著者名 Yamashiki Noriyo, Kawabata Kyoko, Murata Miki, Ikeda Shunichiro, Fujimaki Takako, Suwa Kanehiko, Seki Toshihito, Aramaki Eiji, Naganuma Makoto	4. 巻 7
2. 論文標題 Narrowing the Patient?Physician Gap Based on Self-Reporting and Monthly Hepatologist Feedback for Patients With Alcohol-Related Liver Disease: Interventional Pilot Study Using a Journaling Smartphone App	5. 発行年 2023年
3. 雑誌名 JMIR Formative Research	6. 最初と最後の頁 e44762 ~ e44762
掲載論文のDOI（デジタルオブジェクト識別子） 10.2196/44762	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Eiji Aramaki, Shoko Wakamiya, Shuntaro Yada, Yuta Nakamura	4. 巻 31(1)
2. 論文標題 Natural Language Processing: from Bedside to Everywhere	5. 発行年 2022年
3. 雑誌名 Yearbook of medical informatics	6. 最初と最後の頁 243-253
掲載論文のDOI（デジタルオブジェクト識別子） 10.1055/s-0042-1742510	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

〔学会発表〕 計5件（うち招待講演 3件/うち国際学会 0件）

1. 発表者名 荒牧英治
2. 発表標題 自然言語処理による読影テキスト解析の未来
3. 学会等名 日本医用画像工学会大会(JAMIT2022) シンポジウム（招待講演）
4. 発表年 2022年

1. 発表者名 安藤晶, 宇野裕, 矢田峻太郎, 若宮翔子, 荒牧英治
2. 発表標題 患者状態表現抽出のための看護記録アノテーション
3. 学会等名 人工知能学会全国大会（第36回）（JSAI2022）
4. 発表年 2022年

1. 発表者名 荒牧英治
2. 発表標題 ドメイン言語処理でいかに戦うか：医療言語処理の現在
3. 学会等名 第9回自然言語処理シンポジウム（招待講演）
4. 発表年 2022年

1. 発表者名 荒牧英治
2. 発表標題 NLP-DX： 医療言語処理が加速する医学研究
3. 学会等名 関西製薬協会デジタルヘルスセミナー
4. 発表年 2022年

1. 発表者名 荒牧英治
2. 発表標題 医療言語処理ことはじめ
3. 学会等名 自然言語処理学会チュートリアル（招待講演）
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分 担 者	工藤 紀子 (Noriko Kudo) (30751151)	奈良先端科学技術大学院大学・先端科学技術研究科・博士研究員 (14603)	

6. 研究組織（つづき）

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究分担者	高田 正泰 (Takada Masahiro) (50452363)	京都大学・医学研究科・准教授 (14301)	
研究分担者	川口 展子 (Nobuko Kawaguchi) (60771540)	京都大学・医学研究科・特定助教 (14301)	
研究分担者	山敷 宣代 (Noriyo Yamashiki) (90420215)	関西医科大学・医学部・講師 (34417)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関