

博士学位論文要約 2024 年度

自然言語処理技術を用いた医療記録からの  
症状抽出方法に関する研究  
～患者の主観的情報に焦点を当てて～

慶應義塾大学大学院薬学研究科

大野 由紀子

## 【緒言】

自然言語とは、私たちが日常で用いる言語である。詳細で自由度の高い内容が大量に作成されることから医療分野においても利活用が期待される情報源であるが、構造化されておらず、また意味に曖昧性を含むことから自然言語処理による解析が求められる。自然言語処理技術の一つである固有表現認識 (Named entity recognition: NER) を用いることで、テキストから病名や症状といった臨床情報を収集することができる。医療テキストの中でも、患者が発した生の情報 (Subjective 情報; S 情報) に注目が集まっている。医師は患者の症状を過小評価・過少報告する傾向にあることが知られており、一方で患者の自己報告アンケートの有用性が認められている。また、主観的な症状を最初に気づくのは患者であることから、S 情報をモニタリングすることで有害事象のモニタリングや分析が期待できる。特に、薬剤師は薬学的観点から患者指導を行うため、薬剤師記録中の S 情報に有害事象情報が豊富に含まれる。しかしこれまで、S 情報に注目した研究は主に英語で記述されたテキストを対象としており、日本語を対象とした病名や症状の NER システムは、ルールベースのものや、患者層を限定したデータ、仮想患者のデータを用いたものとどまる。

そこで本研究では、次の 4 つの検討を行った。

第一章: 医師記録を学習済みの病名や症状を収集する自然言語処理システムの薬剤管理指導記録への適用

第二章: 医師記録を学習済みの自然言語処理システムを用いたバンコマイシン塩酸塩による急性アレルギー性皮膚症状の症例検出

第三章: 薬剤師記録中の患者の訴えから病名や症状を抽出する症状抽出システムの性能向上: 学習量増加による効果

第四章: 薬剤師記録中の患者の訴えから病名や症状を抽出する症状抽出システムの性能向上: 他施設データ追加学習による効果

## 第一章: 医師記録を学習済みの病名や症状を収集する自然言語処理システムの薬剤管理指導記録への適用

### 【目的】

医師記録を用いて機械学習 (fine-tuning) した病名抽出システムである MedNER-J を薬剤管理指導記録に適用できるか検証した。

### 【方法】

2018 年 4 月～ 2019 年 3 月に慶應義塾大学病院でセファゾリンナトリウムを投与された患者 49 名、60 件の SOAP 形式の薬剤管理指導記録中の S 情報、Objective (O)

情報、Assessment (A) 情報、Plan (P) 情報に MedNER-J を適用し、病名や症状の NER、抽出した症状の発現の有無を判定する肯定否定判定を行った。研究者は目視で同様の作業を行い、NER の性能と NER および肯定否定判定の性能を、適合率、再現率、および両者の調和平均である F1-score を用いて評価した。

本検討は慶應義塾大学医学部倫理委員会より承認を受けて実施した。（承認番号：20200067「服薬指導記録を用いた副作用に特徴的な語彙の抽出に関する研究」）

#### 【結果】

S、O、A、P 情報でそれぞれ、NER の F1-score は 0.46、0.70、0.76、0.35、肯定否定判定を含めた F1-score は 0.28、0.39、0.64、0.077 であった。

#### 【考察】

O、A 情報からの NER の F1-score はそれぞれ 0.70、0.76 であり、S、P 情報の F1-score (0.46、0.35) よりも高かった。これは O、A 情報に MedNER-J の教師データである症例報告と同様に専門用語の病名・症状が多く含まれることによると考えられた。一方、肯定否定判定を含む F1-score が A 情報のみで高かった (F1-score=0.64) のは、A 情報の記載様式と記載内容が、症例報告と類似することに起因すると考えられた。

MedNER-J は薬剤管理指導記録から NER および肯定否定判定を実施でき、特に A 情報において医師記録の分析に近い性能を示した。一方で、A 情報以外の記録の分析には課題があり、MedNER-J を薬剤管理指導記録に適用するためには、S 情報を中心に教師データを増強する必要があることが判明した。

## 第二章：医師記録を学習済みの自然言語処理システムを用いたバンコマイシン塩酸塩による急性アレルギー性皮膚症状の症例検出

#### 【目的】

医療テキストから病名・症状を収集できる病名抽出システムを活用することにより、有害事象モニタリングの自動化が期待される。そこで、バンコマイシン塩酸塩 (VCM) を投与された患者の看護記録を MedNER-J とルールベースのアルゴリズムを組み合わせることで分析することにより、VCM によるアレルギー性皮膚症状（アレルギー症状）を発現した症例を検出し、発現割合およびその推移を特定できるか調査した。

#### 【方法】

2018 年 4 月 2 日～ 2018 年 6 月 29 日と 2019 年 3 月 25 日～ 2019 年 3 月 29 日の 2 つの期間に、慶應義塾大学病院にて VCM を点滴投与された患者の看護記録を MedNER-J に適用し、発現があった症状を収集した。アレルギー症状に関する論文、ICD-10、ICH 国際医薬用語集 (MedDRA) を参考にアレルギー症状に該当する症状の一

覧（用語リスト）を作成し、看護記録から収集された症状と完全一致または部分一致した症例をアレルギー症状発現症例（アレルギー症例）と見なした。全期間における MedNER-J の看護記録分析性能、MedNER-J とルールベースのアルゴリズムを組み合わせた手法（提案手法）の各期間におけるアレルギー症例検出性能（適合率、再現率、F1-score）、発現割合とその推移を調査した。

本検討は慶應義塾大学医学部倫理委員会および慶應義塾大学薬学部 人を対象とする研究倫理委員会より承認を受けて実施した。（承認番号：20221114「自然言語処理システム MedNER-J を用いた、バンコマイシン塩酸塩によるレッドネック症候群症状の医療文書からの抽出」、承240228-4「自然言語処理システム MedNER-J を用いた、バンコマイシン塩酸塩によるレッドネック症候群症状の医療文書からの抽出」）

#### 【結果】

MedNER-J で看護記録を分析した際の NER 単体、肯定否定判定を含む F1-score はそれぞれ 0.65, 0.60 であった。提案手法による症例検出性能は F1-score = 0.81-0.82 であった。2 つの期間における症状発現割合の変動は提案手法と研究者による目視確認のいずれでも 0.05 パーセントポイントと等しい値であった。

#### 【考察】

MedNER-J による看護記録分析の性能は、先行研究と概ね同等、または先行研究を上回った。さらに、アレルギー症例検出性能は高く（F1-score = 0.82、0.81）、研究者による目視と同等の発現率の変動を検出することができた。以上より、VCM によるアレルギー症状のモニタリングに提案手法が利用できる可能性が示唆された。

### 第三章：薬剤師記録中の患者の訴えから病名や症状を抽出する症状抽出システムの性能向上：学習量増加による効果

#### 【目的】

S 情報を分析対象とした高性能な症状抽出システムを構築するために、システムの学習量を段階的に増やし、これに伴う性能の推移を明らかにした。

#### 【方法】

2018 年度の慶應義塾大学病院における薬剤管理指導記録に含まれる S 情報 6,559 症例、12,004 件に対し前処理と病名・症状のアノテーションを行い、教師データとした。10 段階にかけて教師データを増量しながら 10 分割交差検証により、機械学習モデルである BERT-CRF (bidirectional encoder representations from transformers-conditional random field) を fine-tuning し、適合率、再現率、F1-score を指標として性能を評価した。また、構築したシステムを症例報告に適用して汎用性を評価した。

本検討は慶應義塾大学医学部倫理委員会より承認を受けて実施した。（承認番号：20200067「服薬指導記録を用いた副作用に特徴的な語彙の抽出に関する研究」）

### 【結果】

学習量 1,200 件から 12,004 件にかけて、システムの F1-score は 0.67 から 0.82 へと上昇した。学習量 3,600 件で F1-score は 0.78 となり、以降、学習量に対する性能向上の効率が減少した。性能の向上に伴い、システムの誤抽出によるエラーが最も減少し、適合率の上昇に寄与した。症例報告での検証では、学習量 1,200 件から 12,004 件にかけて、F1-score 0.34 から 0.41 へ増加した。

### 【考察】

学習量 3,600 件以降、学習量の増大に伴う性能の向上には飽和傾向がみられたものの、大量のデータを fine-tuning することにより、薬剤管理指導記録の S 情報を分析可能な高性能な症状抽出システムを構築することができた。

## 第四章：薬剤師記録中の患者の訴えから病名や症状を抽出する症状抽出システムの性能向上：他施設データ追加学習による効果

### 【目的】

これまで、追加の fine-tuning (2nd fine-tuning) が NER を行うシステムの性能に与える純粋な影響は不明であった。そこで、病院の薬剤管理指導記録中の S 情報で fine-tuning したモデルに、薬局薬歴中の S 情報を 2nd fine-tuning し、その効果を明らかにすることで、他施設への応用を踏まえた効率的なシステムの性能向上方法を検討した。

### 【方法】

教師データとして、病院の薬剤管理指導記録に含まれる S 情報 2,794 症例、3,600 件（K 病院データ）および薬局薬歴に含まれる S 情報 2,250 症例、2,400 件（H 薬局データ）を用いた。先行研究にて K 病院データ 3,600 件を fine-tuning した BERT-CRF をベースモデルとし、H 薬局データを増量しながら 10 分割交差検証により 2nd fine-tuning した（追加学習モデル）。加えて、fine-tuning の順序による影響を調べるため、新たに H 薬局データ、K 病院データの順に fine-tuning したモデル（逆順モデル）を構築した。評価指標として適合率、再現率、F1-score を用い、各モデルの K 病院データ、H 薬局データに対する性能を比較した。

本検討は慶應義塾大学医学部倫理委員会および慶應義塾大学薬学部 人を対象とする研究倫理委員会より承認を受けて実施した。（承認番号：20200067「服薬指導記録を用

いた副作用に特徴的な語彙の抽出に関する研究」、承 240228-3「薬局の薬歴データを用いた病名抽出システムの構築および性能向上」)

#### 【結果】

2,400 件の 2nd fine-tuning により、追加学習モデルはベースモデルと比較して、H 薬局データに対する性能がわずかに上昇した一方（追加学習モデルの F1-score = 0.78、ベースモデルの F1-score = 0.74）、K 病院データに対する性能は大幅に低下した（追加学習モデルの F1-score = 0.62、ベースモデルの F1-score = 0.78）。追加学習モデルの H 薬局データに対する性能は、追加学習量が一定量に達するまでベースモデルの性能を下回った。H 薬局データに続いて K 病院データを fine-tuning した逆順モデルでは、ベースモデルに近い性能が得られた（K 病院データに対する F1-score = 0.77、H 薬局データに対する F1-score = 0.76）。

#### 【考察】

K 病院データのみを用いて fine-tuning したベースモデルは H 薬局データを 2nd fine-tuning した追加学習モデルと比較して、双方の施設のデータに対して安定して高い性能であった。このことから、複数施設のデータを混合せず、順番に教師データとして用いる場合、システムの性能は教師データの施設数でなく、最後に用いた教師データの内容と学習量に依存する可能性が示唆された。

#### 【結語】

本研究では、既存の医師記録を教師データとしたシステムの薬剤師記録分析性能を調査し、薬剤師記録分析のベースラインを明らかにした。また、実際に薬剤師記録の S 情報を教師データとして、患者の発話から多様な症状を高性能に抽出するシステムを構築し、教師データの増量と他施設データの 2nd fine-tuning による学習効果を初めて報告した。さらに、既存システムの看護記録からの有害事象抽出への利用可能性を提示した。このことから、本研究で構築した S 情報を教師データとしたシステムを用いて S 情報を分析することで、今後、患者の発話を元にした報告バイアスの少ない有害事象モニタリングの実現が期待される。

#### 【主論文に関する原著論文】

Yukiko Ohno, Riri Kato, Haruki Ishikawa, Tomohiro Nishiyama, Minae Isawa, Mayumi Mochizuki, Eiji Aramaki, Tohru Aomori. Using the Natural Language Processing System Medical Named Entity Recognition-Japanese to Analyze Pharmaceutical Care Records: Natural Language Processing Analysis. *JMIR Formative Research*. 2024; 8: e55798.

Yukiko Ohno, Tohru Aomori, Tomohiro Nishiyama, Riri Kato, Reina Fujiki, Haruki Ishikawa, Keisuke Kiyomiya, Minae Isawa, Mayumi Mochizuki, Eiji Aramaki, Hisakazu Ohtani. Performance Improvement of the Natural Language Processing Tool to Extract Patient Narratives Related to Medical States From Japanese Pharmaceutical Care Records by Increasing the Amount of Training Data: Natural Language Processing Analysis. *JMIR Medical Informatics*. 2025; in press.

【参考論文】

大野 由紀子, 青森 達, 清宮 啓介, 石川 春樹, 西山 智弘, 井澤 美苗, 望月 眞弓, 荒牧 英治, 大谷 壽一. 人工知能学会全国大会論文集. 自然言語処理システムを用いた医療文書中の病名・症状の分析. 2S6-0S-7a-03. 2024.