

# SIGNAL CATCHER: 医学論文を対象とした医薬品有害事象 自動判定システムの構築

氏家 翔吾<sup>1</sup> 岡 守男<sup>2</sup> 島谷 浩司<sup>2</sup> 井阪 航<sup>2</sup> 有本 茜<sup>2</sup>  
矢田 峻太郎<sup>1</sup> 若宮 翔子<sup>1</sup> 荒牧 英治<sup>1</sup>

<sup>1</sup> 奈良先端科学技術大学院大学 <sup>2</sup> 富士ゼロックス株式会社

{ujiie, s-yada, wakamiya, aramaki}@is.naist.jp

{Morio.Oka, shimaya.koji, wataru.isaka, arimoto.akane}@fujixerox.co.jp

## 1 はじめに

近年、医療分野で生成される文書（以下、医療文書）へ自然言語処理技術を応用し、医療現場で有効活用する研究に期待が高まっている。しかし、どのように実際の業務に使うかについては、解析ミスへの対応や説明可能性の不足などの問題から、十分に議論されておらず、医療現場での本格的な実用化には至っていない。

このような背景の中、我々は実用化が近い医療言語処理技術として、医薬品の安全性監視があると考えている。医薬品の安全性監視とは一般には製薬会社が行うもので、各製薬会社は自社薬について論文や診療記録などを対象に、医薬品が投与された結果生じる、好ましくない医療上の出来事（以下、有害事象）の監視を行なっている。この医薬品の安全性監視業務は、監視対象となるデータにあらかじめ大まかな分類や優先度などを人手で付与する一次スクリーニングと、より専門的な判断を要する二次スクリーニングに分けて行うことで、効率化が図られている。しかし、年間数千件の論文を人手で監視する必要がある上に、一次スクリーニングにおいても1論文あたり15分程度要するため、膨大なコストがかかっている。このコスト的な需要に加え、一次スクリーニングでは過誤があっても、二次スクリーニングにて挽回できるため、情報技術を適用しやすいという利点がある。

このため、有害事象を分類・抽出する研究は多くあり、医学論文 [5]、退院サマリなどの電子カルテ [6]、ソーシャルメディアへの投稿 [3] などから有害事象の抽出が試みられてきた。中でも医学論文からの有害事象抽出では大規模なコーパスが作成され [4]、盛んに研究が行われている。Negiらはこのコーパスを用いて有害事象の記述のある文を分類し、その中で疑わしい医薬品を抽出した [9]。しかし、このコーパス [4] は、医

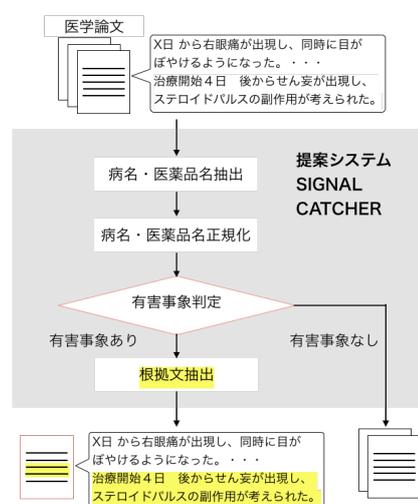


図 1: 提案システムの構成

薬品と症状の関係が1文内に記述されており、かつ明確に因果関係のあるもののみを対象としている。日本では、医薬品と症状の因果が疑わしい関係も監視対象である<sup>1</sup>ため、これらの実験設定では実用化できない。

本研究では、医学論文に対して医薬品との関連が疑われる有害事象の判定を行い、医薬品の安全性監視業務における一次スクリーニングを自動化する実用的なシステムの開発と評価を行う。実際の業務でシステムを使用する際には、システムの出力を手でチェックすることを考慮すると、有害事象判定の根拠となった文（以下、根拠文）も出力し、説明可能性を向上させることが望ましい。そこで、入力論文に対して文書分類を行い、論文単位で有害事象判定をした上で、文分類により根拠文を抽出することで、説明可能性を考慮したシステムを構築した。提案システム SIGNAL CATCHER の構成を図1に示す。

<sup>1</sup><https://www.pmda.go.jp/files/000156127.pdf>

提案システムの根拠文抽出では、日本での監視業務を考慮して、医薬品と症状の関係を1文内に限定せず、明確な因果関係以外にも因果が疑わしいものを扱っている。そのため、先行研究よりも現実的で難しいタスクに取り組み、実用に耐えうるシステムを構築した。

筆者らの知る限り、実用可能なレベルの医療言語処理システムはまれであり、本研究は今後の医療言語処理の導入のさきがけとなる成果であると考えている。

## 2 提案システム

図1に示すように、提案システム SIGNAL CATCHER は、入力論文中の病名・医薬品名を抽出(2.1節)・正規化(2.2節)し、入力論文について有害事象を判定して、根拠文を抽出(2.3節)するパイプラインシステムである。

### 2.1 病名・医薬品名抽出

病名と医薬品名は、有害事象を手でチェックする場合に重要となり、有害事象の自動判定の際にも有用な特徴量になると見込まれる。そのため、入力論文について、病名の抽出に特化した固有表現抽出器 MedEX/J [2] および同様の手法で学習した医薬品名抽出器を用いて、病名と医薬品名を抽出した。

### 2.2 病名・医薬品名正規化

抽出した病名・医薬品名について、表記揺れを吸収し、正規化を行なった。正規化には万病辞書 [7] と百薬辞書<sup>2</sup>を用いた。両辞書は、医学論文や診療録などの電子カルテに記載された病名・医薬品名(以下、出現名)と、標準規格で規定された病名・医薬品名(以下、標準名)が紐づけられている。

抽出した表現に対して、両辞書に記載された出現名との類似度 [1] を計算し、類似度が最も高い出現名に対応する標準名で正規化した。ただし、両辞書に記載されたすべての出現名に対して類似度が閾値(今回は0.3とした)を下回る場合、辞書に該当病名・医薬品名が存在しないものとして標準名は付与しなかった。

### 2.3 有害事象判定と根拠文抽出

#### 2.3.1 概要

本研究では、有害事象判定と根拠文抽出をそれぞれ文書分類、文分類タスクとみなした。近年では、ニュー

ラルネットによる文書分類の際に、文レベルでの注意機構を用いる階層的なモデルが提案されている [11]。しかし、ニューラルネットの学習には大量の学習データが必要である上、注意機構の重みを分類の根拠として扱うのは解釈性に乏しい [8]。そのため、本研究では有害事象判定と独立して文分類することで明示的に根拠文を抽出した。具体的には、入力論文を文書分類し、有害事象判定した上で、文書分類において有害事象に関連すると判定された論文 (Adverse Effect Related Article; AERA) の各文を分類し、根拠文を抽出した。

本システムでは、アノテーションのコストが高く学習データを大量に用意することが困難であるため、ロジスティック回帰を用いて分類した。ロジスティック回帰はパラメータが少なく比較的少量のデータに対しても適用でき、各特徴量の寄与の分析が容易な機械学習手法である。

#### 2.3.2 特徴量

有害事象判定と根拠文抽出では以下の特徴量を使用した。

##### Bag of Words (BoW)

有害事象判定には、「医薬品の投与を中止した。」などの有害事象を示唆する表現が重要な手がかりになると考えられる。そのため、本文の単語を Bag of Words とし、出現頻度を特徴量として用いた。単語の分かち書きには MeCab<sup>3</sup>を使用し、一度しか出現しない単語は除去した。

##### Medical Expression (ME)

病名や医薬品名は有害事象判定に有用であると考えられる。病名・医薬品名には様々な表現があり、表現の差異も考慮するため、病名・医薬品名の出現名と標準名の出現頻度をそれぞれ特徴量として用いた。病名・医薬品名の出現名、標準名については、MedEX/J で抽出し、正規化したものを使用した。また、病名・医薬品名のそれぞれの出現頻度の総和も特徴量として使用した。

##### Context

根拠文抽出において、AERA と判断される根拠文は、複数文にまたがっている場合がある。予備調査では、多くの場合高々2文が連続していることがわかった。そこで、文分類においては前後の文の特徴量も用いた。

<sup>2</sup><http://mednlp.jp> のリソースを参照のこと

<sup>3</sup><https://taku910.github.io/mecab/>

### 3 実験

#### 3.1 実験材料

実験材料として、PDF形式の医学論文509本の概要を光学文字認識技術でテキスト化したものを使用した。これらは日本語で記述された症例報告（個々の患者の症状や診断・治療の詳細をまとめた論文）などで構成されている。

論文中の各文に対して、有害事象判定の根拠となる文に正解ラベル「T」を、それ以外の文に「F」のラベルを付与し、論文中の文のうち1文でも正解ラベルが付与された論文をAERAとした。509論文のうち300論文がAERAである。

#### 3.2 実験設定

層化5分割交差検定により、モデルの学習と評価を行なった。文分類では、文書単位で学習データとテストデータの分割を行った。また、AERAでない論文中の文には正解ラベルが付与されていないため、文分類で学習データの全論文を用いると、各ラベルのデータ量に大きな不均衡が生じる。これを避けるため、文分類では、学習データのうちAERAの論文のみで学習し、テストデータの全論文を評価に用いた。

評価にはPrecision-Recall曲線を用いた。Precision-Recall曲線は閾値を変化させた際の再現率と適合率を可視化したものであり、医薬品の安全性監視のような再現率と適合率のトレードオフが重要なタスクの評価に適している。

文分類および文書分類に使用したロジスティック回帰には、scikit-learn [10] のLogisticRegressionクラスを用いた。パラメータはデフォルトのものを使用した。

#### 3.3 実験結果

2.3.2節で述べた特徴量を用いて実験を行った結果を示す。有害事象判定のPrecision-Recall曲線を図2aに、根拠文抽出のPrecision-Recall曲線を図2bに示す。交差検定の平均として、有害事象判定ではF値0.914という実用性能を達成した一方で、根拠文抽出では0.439という低いF値にとどまっている。

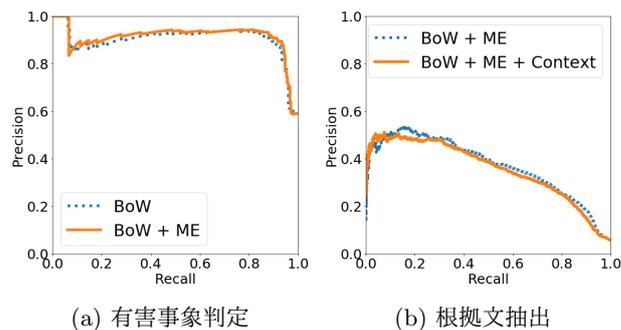


図 2: Precision-Recall 曲線

### 4 考察

#### 4.1 誤り分析

全ての特徴量を用いた根拠文抽出モデルの分類誤りを分析した。表1は、システムの分類例である。各例は、最初の句点あるいは全角ピリオドまでが前文、以降が該当文である。なお表中のT、Fはそれぞれ該当文のラベルを表している。本節ではシステムが誤分類した(c)から(f)について考察する。

(c)、(d)は、根拠文抽出に前文の情報を必要とし、分類を誤ったと考えられる例である。各例では、該当文のみを考えた場合は報告の根拠にならない。しかし、該当文で言及された症状が前文で言及された医薬品による有害事象である可能性を否定できないため、報告の根拠にすべきである。分類の際、前後の文の特徴量を追加しているにも関わらず分類誤りが生じている。この原因について次節の特徴量分析において考察する。

(e)、(f)は、一般論についての言及を実際の症例と誤って分類したと考えられる例である。各例の2文目には、医薬品を原因とする症状が一般論として記述されている。一般論と実際の症例は表現上似通っており、誤りにつながったとみられる。

#### 4.2 特徴量分析

該当文のみの特徴量を用いた場合と前後の文の特徴量を用いた場合の比較を行なった。図2bに示すように、前後の文の特徴量を用いた場合、精度が悪化した。ロジスティック回帰の回帰係数で特徴量の寄与を分析したところ、「中止」「抑制」の順に寄与が大きかった。一方、前後の文の特徴量で、分類に寄与した特徴量の上位5例には後の文の「同日」のみが入った。このことから、前後の文の特徴量を追加することで、特徴量の次元が学習データに対して大きくなり、その多くがノイズとして作用したと考えられる。

表 1: 根拠文抽出における誤り分析

	正解	予測	入力文
(a)	T	T	MTX <sup>i</sup> アダリムマブ投与開始。MTX 副作用で継続困難となりセルトリズマブベゴルに変更。
(b)	T	T	【症例】74 歳、男性【既往歴】74 歳：ABPC / SBT <sup>ii</sup> で発疹。CEZ <sup>iii</sup> でアナフィラキシーショック。
(c)	T	F	59 歳時より RA <sup>iv</sup> に対して MTX 内服。2018 年 3 月に呼吸困難があり受診。
(d)	T	F	【症例】79 歳、女性【主訴】【現病歴】巨細胞性動脈炎でプレドニゾン 40mg / 日とメトトレキサート 8mg / 週を内服し、糖尿病でインスリンを使用していた。2 週間前から倦怠感・食思不振が続き入院となった。
(e)	F	T	スティーヴンス・ジョンソン症候群 (Stevens-Johnson syndrome : SJS) は、発熱と眼、口唇、外陰部などの皮膚粘膜移行部における重度の粘膜疹、紅斑と表皮の壊死性傷害に基づく水疱・びらんを特徴とする。多くは医薬品が原因で発症する最重症型薬疹の一つと考えられるが、一部はウイルスやマイコプラズマなどの感染に伴って発症する。
(f)	F	T	図 10 臨床経過。Reporting System の報告によると過去 5 年間の中で TdP <sup>v</sup> の原因になった薬剤の約 25% (92 / 372 件) はニューキノロン系の抗菌薬であり、その多くはレボフロキサシンによるものであった 2)。
(g)	F	F	【症例】高血圧の既往がある 70 歳男性。X 日から右眼痛が出現し、同時に目がぼやけるようになった
(h)	F	F	【症例】79 歳女性。中血管炎の診断で半年前からプレドニゾン 60 mg とメトトレキサート 6 mg を内服中であった。

<sup>i</sup> メトトレキサート (methotrexate) の略

<sup>ii</sup> アンピシリン/スルバクタム (ampicillin/sulbactam) の略

<sup>iii</sup> セファゾリン (cefazolin) の略

<sup>iv</sup> 関節リウマチ (Rheumatoid Arthritis) の略

<sup>v</sup> Torsades de Pointes の略で不整脈の一種

## 5 おわりに

本研究では、医薬品の安全性監視における一次スクリーニングを自動化する実用的システムを開発した。医薬品による有害事象に関連する論文に対して実用精度でスクリーニングが行えることを示し、人手精査前のスクリーニングに自然言語処理技術が適用できることを示唆した。今後は、本技術の実用化を進めたい。最後に、スクリーニングの根拠文抽出は前後の文脈を十分に捉えられておらず、素性やモデルの設計に改善の余地がある。根拠文抽出として前後の文脈をより高度に考慮する分類モデルを検討したい。

## 参考文献

- [1] Eiji Aramaki, Takeshi Imai, Kengo Miyao, and Kazuhiko Ohe. Orthographic disambiguation incorporating transliterated probability. In *IJCNLP 2008*, 2008.
- [2] Eiji Aramaki, Ken Yano, and Shoko Wakamiya. MedEx/J: A one-scan simple and fast NLP tool for japanese clinical texts. *MEDINFO 2017*, Vol. 245, p. 285, 2018.
- [3] Xiaoyi Chen, Carole Faviez, Stéphane Schuck, Lillo-Le Louët, Nathalie Texier, Badisse Dahamna, Charles Huot, Pierre Foulquié, Suzanne Pereira, Vincent Leroux, et al. Mining patients' narratives in social media for pharmacovigilance: adverse effects and misuse of methylphenidate. *Frontiers in Pharmacology*, Vol. 9, p. 541, 2018.
- [4] Harsha Gurulingappa, Abdul Mateen-Rajpu, Angus Roberts, Juliane Fluck, Martin Hofmann-Apitius, and Luca Toldo. Development of a benchmark corpus to support the automatic extraction of drug-related adverse effects from medical case reports. *Journal of Biomedical Informatics*, Vol. 45, No. 5, pp. 885–892, 2012.
- [5] Harsha Gurulingappa, Abdul Mateen-Rajpu, and Luca Toldo. Extraction of potential adverse drug events from medical case reports. *Journal of Biomedical Semantics*, Vol. 3, No. 1, p. 15, 2012.
- [6] Sam Henry, Kevin Buchan, Michele Filannino, Amber Stubbs, and Ozlem Uzuner. 2018 n2c2 shared task on adverse drug events and medication extraction in electronic health records. *Journal of the American Medical Informatics Association*, Vol. 27, No. 1, pp. 3–12, 2019.
- [7] Kaoru Ito, Hiroyuki Nagai, Taro Okahisa, Shoko Wakamiya, Tomohide Iwao, and Eiji Aramaki. J-MeDic: A Japanese Disease Name Dictionary based on real clinical usage. In *LREC 2018*, 2018.
- [8] Sarthak Jain and Byron C. Wallace. Attention is not Explanation. In *NAACL 2019*, pp. 3543–3556, June 2019.
- [9] Kajal Negi, Arun Pavuri, Ladle Patel, and Chirag Jain. A novel method for drug-adverse event extraction using machine learning. *Informatics in Medicine Unlocked*, p. 100190, 2019.
- [10] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, Vol. 12, pp. 2825–2830, 2011.
- [11] Zichao Yang, Diyi Yang, Chris Dyer, Xiaodong He, Alex Smola, and Eduard Hovy. Hierarchical attention networks for document classification. In *NAACL 2016*, pp. 1480–1489, 2016.