

令和元年度 厚生労働科学研究費 補助金
(政策科学総合(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

研究代表者 今任 拓也 国立がん研究センター・社会と健康研究センター・主任研究員

研究要旨:【目的】本研究は、診断基準が比較的明確であり、致死性の高い副作用であるステイブンス・ジョンソン症候群(SJS)および中毒性表皮壊死融解症(TEN)を対象とし、PMDAにおいて管理されている副作用個別症例報告を用いて、近年、医療を始めとした様々な分野で注目されている人工知能技術を応用し、副作用判定評価システムの構築のための試行的評価を実施した。前年度は、副作用判定に人工知能を応用するための基盤整備に先行し、まず、人手により素性を付与した副作用判定の試行的評価を実施したため、今年度は、主に基盤整備を実施し、改めて、全行程を通じた副作用判定の試行的評価を行うこととした。【方法】PMDAにて管理されている副作用個別症例報告書(SJS:1410件、TEN:200件)のテキストデータを用いた。素性抽出のためのトレーニングデータの作成には、brat(brat rapid annotation tool)を用い、アノテーション作業を行った。前年度の副作用自動判定の評価に用いたSJSの副作用個別症例報告書200件はテスト用データとして用い、残りの1410件のうち、981件を学習用データ、429件をバリデーションデータとして用いた。素性アノテーション用学習モデルには、SJSに関連する論文を含む広範囲にわたる医学雑誌の論文、JAPICが提供する医薬品添付文書および独自開発の自動検索装置によりWEBから収集された医療分野のテキストからなる総計6千万文の日本語テキストを用いてBert用pretrainedモデルを構築して用いた。最後に、全行程を通じた副作用判定の精度評価は、前年度と同様に、素性の評価が容易なMaximum Entropy Classifier(MEC)を用いて5分割交差検証により行った。【結果】自動アノテーションについては、0-tagを含むすべてのタグ付けのテストデータの精度は96.5%、0-tagを除いた精度は69.4%であった。また、この自動アノテーションの結果をもとに副作用自動判定を行った結果、81.5%の判定精度が得られた。【考察】前年度に実施した素性を人手で付与した場合の副作用自動判定の精度は86.0%であったのに対し、今年度を実施した自動アノテーションの結果をもとにした副作用自動判定の結果では、4.5%低下したが、これは当初の目標としていた80%の精度を上回る結果となった。今回得られた結果は、安心して使用できるレベルではないと考えられ、さらに精度を上げるためには、今回の研究では捉えきれない特徴を考慮すること、また、副作用個別症例報告書の記載内容の記載者によるぶれを減らすなど、報告書の形式についても検討していく必要があると考えられる。また、副作用自動判定システムの構築には、それぞれの副作用の判定に適した特徴量の抽出が今後の課題であると考えられる。

研究代表者:

今任拓也・国立がん研究センター・社会と健康研究センター・主任研究員

研究分担者:

潮田 明・産業技術総合研究所・人工知能研究センター・招聘研究員

斎藤嘉朗・国立医薬品食品衛生研究所・医薬安全科学部・部長

森谷純治・医薬品医療機器総合機構・医薬品安全対策第1部・主任専門員

相原道子・横浜市立大学・耳鼻科・教授

A. 研究目的

本研究は、近年、医療を始めとした様々な分野で注目されている人工知能技術を応用し、年々増加傾向にある副作用症例報告をより効果的かつ適切に副作用を評価する技術の開発を目標としている。

そこで、診断基準が比較的明確であり、致死性の高い副作用であるスティーブンス・ジョンソン症候群(SJS)および中毒性表皮壊死融解症(TEN)を対象とし、PMDAにおいて管理されている副作用個別症例報告を用いて、人工知能を応用し、副作用判定評価システムの構築のための検討をこれまで行ってきた。本研究の計画は、

副作用報告の評価に人工知能を応用するための基盤整備と人工知能を用いた試行的な副作用判定の評価という、2つの工程に分けて検討することとしている。

当初の計画では、前年度より基盤整備を開始し、今年度は主に の試行的な副作用判定の評価を実施する予定としていたが、前年度に個人情報保護などの観点により実務的な制約が生じたため、先行して比較的個人情報保護の影響が少ない、人工知能を用いた副作用判定評価を行った。よって、本年度は、主に基盤整備を実

施し、改めて、全行程を通じた副作用の試行的評価を行うこととした。

B. 研究方法

本研究の概要については、図1に示し、本年度実施した部分については、赤枠で示した。

対象副作用

診断基準が比較的明確であり、致死性の高い副作用であるスティーブンス・ジョンソン症候群(SJS)および中毒性表皮壊死融解症(TEN)を対象副作用とした。

使用データ

PMDAにて管理されている副作用症例個別報告書(SJS:1410件、TEN:200件)の経過欄のテキストデータを主に用いた。

素性抽出のためのトレーニングデータの作成

トレーニングデータは、形態素解析を施した副作用症例報告書のテキストに対して、副作用に特徴的な素性(特徴量)のアノテーションを施すことにより作成した。アノテーションには、brat(brat rapid annotation tool)を用いた。

アノテーション作業は、事前にトレーニングされた臨床経験のある看護師が行った。まず、副作用判定において重要な用語(主に単語)の、素性に応じた色付けを行い、次にその結果を参照しながら、形態素解析を施したテキストに対して形態素ごとにIOB2(Inside-outside-beginning)フォーマットを用いた素性のアノテーションを行うといった2段階に分けて実施した。

前年度の副作用自動判定の評価に用いた SJS の副作用症例報告書 200 件についてはテスト用データとして用い、残りの 1410 件のうち、981 件を学習用データとして、429 件をバリデーションデータとして用いた。

トレーニングデータを用いた素性抽出のモデル構築

素性アノテーション用学習モデルには Bert を用いた。アノテーション用トレーニングデータの他に Bert 用の日本語 pretrained モデルが必要であるため、SJS に関連する論文を含む広範囲にわたる医学雑誌の論文、JAPIC が提供する医薬品添付文書および独自開発の自動検索装置により WEB から収集された医療分野のテキストからなる総計 6 千万文の日本語テキストを用いて Bert 用 pretrained モデルを構築した。

副作用判定の精度評価

副作用判定には、前年度と同様に素性の評価が容易な Maximum Entropy Classifier(MEC)を用いて 5 分割交差検証により行った。

C. 結果

自動アノテーションについては、0-tag を含むすべてのタグ付けのテストデータにおける精度は 96.5%、0-tag を除いた精度は 69.4%であった。また、この自動アノテーションの結果をもとに副作用自動判定を行った結果、81.5%の判定精度が得られた。

D. 考察

本研究は、PMDA にて管理されている副作用症例報告の経過欄の情報を基に、人工知能を用

いて、副作用を精度良く判別できるシステムの構築のための基盤整備および、現在人手で行われている副作用判定の人工知能を応用した試行的な自動判定の評価を行うものである。

前年度に実施した素性を人手で付与した場合の副作用自動判定の精度は 86.0%であったのに対し、今年度を実施した自動アノテーションの結果をもとにした副作用自動判定の結果は、4.5%低下したが、81.5%の判定精度が得られた。これは当初の目標としていた 80%の精度を上回る結果となっている。

現在、ディープラーニングを始めとした人工知能に関する研究に大きな期待が寄せられている。特にディープラーニングは特徴量を自ら作り出すことができるようになった点で画期的な技術ではあるが、大規模かつ質の高いデータが必要であり、単独であらゆる分野のさまざまな種類の問題を解決できるような万能な人工知能を生み出すまでには至っていない。したがって、医療の分野も含め、まずは本研究のように、それぞれの分野における具体的な問題に対応できる人工知能を個々に実用化していくための研究が重要であると考えられる。

本研究にて、今回得られた結果は、安心して使えるレベルではないと考えられる。今回は、実際に副作用判定を行った経験のある専門員へのヒアリング、及び重篤副作用疾患別対応マニュアルを基に特徴量を選出したが、副作用判定に重要と考えられる特徴量を完全には捉えきれていないことも考えられる。特に、病状などの記載が具体的かどうかや経過を時系列で追えるかどうかといった記載者に起因する「書きぶり」などといった要因なども結果に大きく影響している可能性が考えられ、使用する情報の質の統一も重要であると考えられる。今回使用し

た副作用個別症例報告のうち、422 症例について、DLST および発熱といった副作用判定に必要な特徴量の記載場所について検討したところ、発熱については、およそ 9 割が経過欄内に記載されているが、DLST の結果については、5 割程度しか経過欄内に記載されていないことがわかった。残りの 5 割は、担当医のコメント欄、検査値欄内など、記載者により異なる項目に記載されていることがわかった(表 1)。今後、機械学習などの人工知能を用いて、このような症例報告などのテキストデータを基に副作用などを判定するためには、記載する副作用症例報告書の項目や記載者によるぶれがないような形式にする必要があると考える。また、副作用判定は、現在人手によって行われている。今回の研究においても、3 人の副作用判定の経験者である PMDA の専門員の結果に基づいて行った。3 人の専門員による結果についても、高い一致率は得られておらず、限られた情報だけでは、人においても副作用判定は難しいことが推測される。よって、人工知能が自分自身の判定結果に付与する確率を活用するなど、本研究によって得られた結果が、今後の機械学習を用いた副作用自動判定システム構築のための手がかりの 1 つとなることを期待する。

本研究計画では、当初、予定していなかった後半部分に当たる副作用自動判定の試行的評価を前年度に実施し、ある程度精度の高い学習モデルの構築ができたため、今年度はテキストデータから副作用判定に必要な特徴量の自動抽出という重要な基盤整備に注力することが可能となった。このことにより作業の効率化が図れたと考えられる。また、今回は、検査値などを使用することなく、比較的診断基準が明確である SJS と TEN という副作用を対象とした。

本研究で使用した技術は、比較的汎用性が高いと考えられるが、それぞれの副作用の判定に適した特徴量の抽出が今後の課題であると考えられる。

E. 倫理審査状況

本研究で使用するデータは、PMDA が所有する既存の副作用個別症例データの二次利用であり、患者のインフォームド・コンセントは取得していない。しかしながら、医薬品の副作用報告は、「医薬品、医療機器等の品質、有効性及び安全性の確保等に関する法律施行規則」第 228 条の 20 第 1 項に基づき実施されるものであり、副作用個別症例情報には、患者氏名、生年月日等などの個人に関する情報は記載されておらず、PMDA にて対応表は保有してないため、特定の個人を識別することは不可能である。また、「人を対象とする医学系研究に関する倫理指針」に則り、PMDA にて承認を受け、国立医薬品食品衛生研究所および産業総合研究所の倫理審査においても承認を受けている。

F. 健康危険情報

該当なし

F. 研究成果

1. 論文発表

該当なし

2. 学会発表

該当なし

①副作用報告の評価に人工知能を応用するための基盤整備



②人工知能を用いた副作用の試行的評価

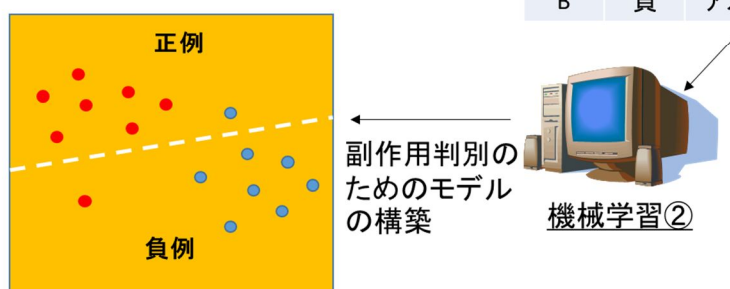


図1 研究の流れ

赤枠内が今年度実施した課題である。

表 1.422 症例中の DLST および発熱情報とその記載場所

	DLST	%	発熱	%
	171	40.52 (422 症例中)	179	42.42 (422 症例中)
経過欄	90	52.63	161	89.94
担当医のコメント欄	90	52.63	11	6.15
検査値欄	102	59.65	62	34.64
経過欄以外	81	47.37	18	10.06