

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))
総合研究報告書

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】医薬品の安全対策に対し、リスクマネジメントプランが導入され、開発段階から市販後までの安全対策の充実強化が図られている。副作用報告は、主に製薬企業により症例報告書の形で医薬品医療機器総合機構(PMDA)に提出されることとなっている。2012年には国内症例4万件程度だったものが2018年には約6万件となっており、年々増加傾向にある。そのため、PMDAでは膨大な症例報告を効率的かつ適切に評価することが必要となっている。そこで、PMDAにおいて管理されている副作用個別症例報告を用い、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、副作用判定に人工知能を応用するための基盤整備を実施し、人工知能を応用したSJSおよびTENの副作用判定の自動化システムの構築を目標とした。本研究は、そのための試行的評価を実施し、その精度について検討を行った。【方法】本研究は、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副作用の試行的評価の2つの工程で実施することとした。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では、SJS、TENの副作用判定において重要と考えられる素性抽出のためのトレーニングデータの作成およびBertを用い、作成したトレーニングデータを用いた素性抽出のモデル構築を行った。また、人工知能を用いた副作用判定の試行的評価には、副作用評価のための機械学習用トレーニングデータを作成し、Maximum Entropy Classifier(MEC)を機械学習器として用い、副作用判定の試行的な評価を実施した。【結果】副作用判定の評価に人工知能を応用するための基盤整備における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、人工知能を用いた副作用の試行的評価においては、81.5%の判定精度が得られた。【考察】3名の専門員の判定精度の概算値が80~90%であったことから、専門家の精度領域に近づいたと言える。しかしながら、現場で人工知能を活用しようとした場合、81.5%の精度ではまだ安心して使えるレベルとは言えない。本研究結果から人工知能を現場で役立てていくための方法の1つとして、人工知能の確信度、すなわち、「人工知能が自分自身の判定結果に付与する確率」の活用が考えられる。また、モデルのエラー分析は専門家の判断基準の明確化に活用できる可能性も期待できる。一方で、副作用個別症例報告の書きぶりも副作用判定に影響する要素の1つと考えられ、副作用判定に人工知能を応用するためには副作用個別症例報告書の形式についても、検討が必要である可能性も示唆された。開発した素性抽出のためのトレーニングデータおよび副作用評価のための機械学習用トレーニングデータはいずれもSJS、TENに特化したものであり、トレーニングデータのクリーニングなどの課題は残っているものの、機械学習用のモデルおよび自然言語

処理の要素技術は他の副作用にも適用可能であり、今後のさらなる拡充が期待される。

研究代表者:

今任拓也・国立医薬品食品衛生研究所・医薬安全科学部・主任研究官

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・医薬品安全対策第一部・主任専門員

相原道子・横浜市立大学・耳鼻科・教授

を効率的かつ適切に評価することが必要となっている。

人工知能は、ヒトの知的な処理を機械によって行う技術であり、ヒトの意思決定に代わるものとなると期待されている。人工知能の研究は、1950年代から続いており (Buchanan, Bruce G. A (Very) Brief History of Artificial Intelligence, AI Magazine(Winter) 2005; 53-60), 2006年ごろにディープラーニングという新たな技術が提唱され、現在では、第三次ブームとして、医療分野を含めた様々な分野において大きな期待が寄せられている。

そこで、本研究は、PMDAにおいて管理されている副作用個別症例報告を用い、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、副作用判定に人工知能を応用するための基盤整備を実施するとともに、人工知能を用いて、SJS および TEN の副作用判定の自動化システムの構築を目的とした試行的評価を実施し、その精度について検討を行った。

A. 研究目的

医薬品の安全対策に対し、リスクマネジメントプランが導入され、開発段階から市販後までの安全対策の充実強化が図られている。一方で、国際共同治験の推進により臨床試験における日本人症例の割合は減少し、日本人における医薬品の開発段階の安全性に関するエビデンスは十分とは言えない。また2017年10月より運用が開始された条件付き早期承認制度により、これまで限られた患者数での安全性検討により発売される場合もあり、市販後における医薬品安全性監視は、益々重要となってきている。

市販後医薬品安全対策の骨格を成すものは、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第68条の10第2項に規定されている副作用等の報告である。本報告は、主に製薬企業により症例報告書の形で直接医薬品医療機器総合機構 (PMDA) に提出されることとなっている。この企業から副作用報告は2012年には国内症例4万件程度だったものが2018年には約6万件となっており、年々増加傾向にある。そのため、PMDAでは膨大な症例報告

B. 研究方法および研究結果

対象副作用

使用するデータの検討から検査値データの使用は、個人情報保護法において要配慮個人情報に該当し、PMDAのセキュリティポリシーでは、外部施設への提供は困難であるとの見解から、検査値情報の使用が困難であったため、診断基準において、検査値情報の必要性が低く、診断基準が明確であるSJSおよびTENを対象副作用とすることとした。

使用データ

主に企業から報告されている副作用個別症例報告の情報をを用いた。使用したデータの項目については、表 1 に示している。正負判定には、PMDA の専門員による 1 次評価の結果を用いた。PMDA における評価において副作用として認められた報告書は正例、副作用として認めるのに十分な所見あるいは情報がないと判断された報告書は負例とした。また実際に PMDA において副作用評価を行っている複数の専門家からのヒアリング結果および「重篤副作用疾患別対応マニュアル」をもとに表 2 に示す素性を副作用判定用の素性として選択した。最終的には SJS 1420 件および TEN 200 件のデータを使用した。なお、研究 2 年目に実施した素性を人手により付与した副作用の試行的評価には、正例 100 件と負例 100 件を使用した。

なお、本研究で使用する副作用個別症例報告は原則非開示情報（機密性情報 3）とされており、これまで情報の二次利用および外部施設へのデータの提供が考慮されていなかったため、個人情報保護の観点を含め、今後のデータの取扱について十分な検討を行い、研究班内でデータの取扱に関する規定を作成した。

本研究の概要については、図 1 に示している。本研究は、

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

人工知能を用いた副作用の試行的評価の 2 工程で実施することとし、3 年計画とした。

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

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。

人工知能を用いた副作用の試行的評価

当初の計画では、人工知能を応用するための基盤整備を実施した後に、本工程を実施する予定としていたが、当初予定していた研究2年目の基盤整備を実施する際に、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、基盤整備に先行して、副作用判定に重要と考えられる素性を人手により付与した場合の副作用自動判定の試行的評価を行った。

副作用評価のための機械学習用トレーニングデータの作成

まず、複数の専門家からのヒアリング結果、および「重篤副作用疾患別対応マニュアル」をもとに表2に示す素性を機械学習用トレーニングデータの素性とした。また、トレーニングデータの作成には、構築したBert用 pretrained モデルを用いた。なお、人手により素性を付与したトレーニングデータを作成した場合は、医学的知識がある者が実施した。

副作用の試行的評価

副作用判定は正例と負例を判別する2値分類であるため、高精度が期待でき、更に素性の評価が容易な Maximum Entropy Classifier (MEC) を機械学習器として使用した。

また、研究2年目に実施した人手により素性を付与した場合の試行的副作用評価の結果より、判定の難しいケースでは評価者間に相当量の判定の揺らぎがあることが明らかになったため、正負データ200件につき、専門員3名による正負判定の再評価を行った。専門員3名による判定は個別に行い、最終判定を機械的に3名の判定結果の多数決により決定し、その判定結果を gold standard (GS) とした。

C. 研究結果

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

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(表2)。

人工知能を用いた副作用の試行的評価

二年目に先行して実施した、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(データは示していない)。一方、最終年度に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、81.5%の判定精度が得られた(表2)

C. 考察

本研究は、PMDAにて整理・管理されている副作用症例報告の経過欄の情報を基に、現在、人的作業により行われている副作用判定に人工知能を応用し、副作用を精度良く、かつ効率的に判別できるシステムの構築を目標とし、そのための試行的評価を行うものである。その結果、当初目標の80%を上回る81.5%の判定精度を達成したが、現場で人工知能を活用しようとした場合、まだ安心して使えるレベルとは言えない。人工知能を現場で役立てて行くための方法の1つとして、人工知能の確信度、すなわち「人工知能が自分自身の判定結果に付与する確率」の活用が考えられる。全200例を学習データとして用いたMECで学習されたモデルにおけるそれぞれの素性に対応する重みを比較し、どの素性が副作用判定において重要であるかを見てみると、皮膚粘膜移行部の広範かつ重度な粘膜病変が最も重要とされ、次に報告書テキスト外

テーブルより得られた被疑薬投与後の症状の発現と判断、その次に皮膚の汎発性の紅斑に伴う表皮の壊死性障害に基づくびらん・水疱となっており、SJS の診断基準と同様に、粘膜病変や壊死性障害に基づくびらん・水疱といった症状に関する記述の有無が副作用判定において重要であることが示唆された。一方で、「発熱」の記載が認められた症例は比較的少なく、「発熱」は SJS の診断において重要な判断基準とされているが、副作用の判断基準としてはやや弱いと考えられた。殆どの症例で発熱することが考えられ、特筆すべき症状としては記載されていないのかもしれない。このように副作用判定のために学習されたモデルが正解判定およびモデルによる判定双方に与える確率を比較することやモデルからみて信頼度の低い正解判定を調べてみると、信頼度が特に低かった事例のなかには、専門家から見てモデルの判断でも間違いとは言えない事例が散見されるなど、このようなモデルのエラー分析は専門家の判断基準の明確化に活用できる可能性も期待できる。

また、今回選ばれた素性をほとんど満たしているにもかかわらず、負例と判断されている症例もあり、まだ捉えきれていない特徴量が存在している可能性も示唆されており、例えば、どのように記載されているのかという要因も重要であることが示唆された。特に病状の経過が時系列で追えるケースや薬剤投与量と投与期間が明確に記載されているケースは信頼性が高く、症例報告書中に主要な文言は揃っているものの、病状の記述に具体性が欠けているために信頼性が低いと判断されていることもわかった。データは示していないが、DLST などは、経過欄、担当医のコメント欄、検査欄など、記載

者によって、記載場所が異なっていることなどから、これらを踏まえ、将来的に人工知能をこのような副作用判定に応用するためには、現在使用している副作用個別症例報告書の形式なども検討する必要があると考えられる。

また、当初、本研究計画では、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、機械学習に有効な特徴量の洗い出しをある程度行ったことによって、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。しかしながら、人工知能によるアノテーションの際に、フォーマットエラーを起こしているものも存在しており、素性アノテーション用トレーニングデータのクリーニングについても今後の検討課題として挙げられる。

本研究は、症例報告書からの副作用に特徴的な素性（特徴量）の自動抽出の枠組みを構築し、開発した素性・素性値ペアのセットから副作用の自動判定を行うモジュールと統合して、全行程を通じて人工知能により副作用判定を行うことのできるモデルを構築し、判定精度の試行的評価を行った。その結果、当初目標の 80% を上回る 81.5% の判定精度を達成した。3 名の専門員の判定精度の概算値が 80% ~ 90% であることから、専門家の精度領域に近づいたと言える。開発した素性抽出のためのトレーニングデータおよび副作用評価のための機械学習用トレーニングデータはいずれも SJS、TEN に特化したものであるが、機械学習用のモデルおよび NLP の要素技術は他の副作用にも適用可能であり、今後の拡充が期待される。

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技術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

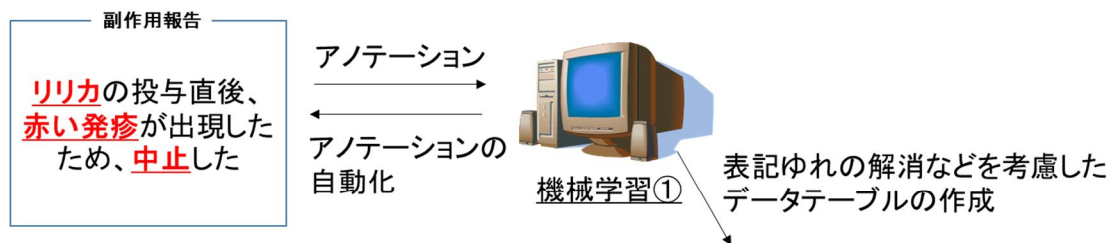
1. 論文発表

該当なし

2. 学会発表

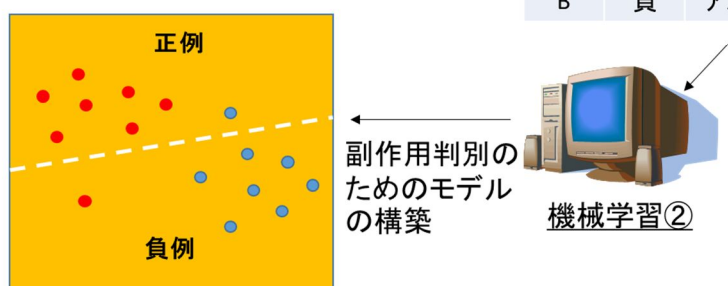
潮田明、今任拓也、森谷純治、斎藤嘉朗、松永雄亮、沼生智晴、見田活、阿川英之、関口遼：
人工知能を活用した副作用症例報告書の試行的評価．第 5 回日本医療安全学会学術総会（2019.2）

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



症例	判定	被疑薬	診断基準 ¹	中止	時間的整合性
A	正	プレガバリン	あり	あり	あり
B	負	アスピリン	なし	なし	なし

②人工知能を用いた副作用の試行的評価



3

図1 研究の流れ

表1 副作用個別症例報告の使用データについて

変数	使用可能状況
身長・体重	使用不可
日付データ（投与日、経過欄日付、過去治療歴日付、副作用発症日） （ただし、被疑薬投与日を起点とした日数とする）	使用可
経過欄（テキストデータ） （ただし、患者名や病院名、検査値などはマスキングする）	使用可
検査値	使用不可
被疑薬	使用可
DLSTの結果（数値の使用は不可、陰性および陽性を使用）	使用可
発熱（数値の使用は不可、発熱は37.5 以上とした）	使用可

表2 使用した素性セット

素性番号	素性	素性値	判断基準
1	皮膚粘膜移行部の広範かつ 重度な粘膜病変	2	有り
		1	有り(やや軽度)
		0	なし
2	皮膚の汎発性の紅斑に伴う表皮の 壊死性障害に基づくびらん・水疱	2	有り
		1	有り(やや軽度)
		0	なし
3	発熱	1	有り
		0	なし
4	病理組織所見	1	有り
		0	それ以外
5	皮膚科によるSJS診断	1	有り
		0	なし
6	被疑薬のDSLTL検査	1	検査有り(陽性)
		0	検査有り(陰性)
		-	検査なし
7	併用薬のDSLTL検査	1	検査有り(陽性)
		0	検査有り(陰性)
		-	検査なし
8	報告書テキスト内に被疑薬投与後の症状発現が記載	1	有り
		0	なし
9	報告書テキスト外テーブルより 被疑薬投与後の症状発現と判断	1	できる
		0	できない

表3 自動アノテーションおよび副作用自動判定の精度

素性のアノテーション	テストデータの アノテーション精度 (%)
全形態素の精度	96.5
O-tag 付与語を除く全形態素の精度	69.4

素性のアノテーション	テストデータの副作用自動 判定の精度 (%)
人手によるアノテーション	86.0
AIによるアノテーション	81.5

参考 SJS の診断基準

主要所見 (必須)	<ol style="list-style-type: none">1 皮膚粘膜移行部(眼,口唇,外陰部など)の広範囲で重篤な粘膜病変(出血・血痂を伴うびらん等)がみられる。2 皮膚の汎発性の紅斑に伴って表皮の壊死性障害に基づくびらん・水疱を認め,軽快後には痂皮,膜様落がみられる。その面積は体表面積の10%未満である。但し,外力を加えると表皮が容易に剥離すると思われる部位はこの面積に含まれる。3 発熱がある4 病理組織学的に表皮の壊死性変化を認める*5 多形紅斑重症型(erythema multiforme [EM] major)**を除外できる。
副所見	<ol style="list-style-type: none">1 紅斑は顔面,頸部,体幹優位に全身性に分布する。紅斑は隆起せず,中央が暗紅色の flat atypical targets を示し,融合傾向を認める。2 皮膚粘膜移行部の粘膜病変を伴う。眼病変では偽膜形成と眼表面上皮欠損のどちらかあるいは両方を伴う両眼性の急性結膜炎がみられる。3 全身症状として他覚的に重症感,自覚的には倦怠感を伴う。口腔内の疼痛や咽頭痛のため,種々の程度に摂食障害を伴う。4 自己免疫性水疱症を除外できる。

*病理組織学的に完成した病像では表皮の全層性壊死を呈するが,少なくとも 200 倍視野で 10 個以上の表皮細胞(壊)死を確認することが望ましい。

**多形紅斑重症型(erythema multiforme [EM] major)とは比較的軽度の粘膜病変を伴う多形紅斑をいう。皮疹は四肢優位に分布し,全身症状としてしばしば発熱を伴うが,重症感は乏しい。SJS とは別疾患である。

参考 TEN の診断基準

主要所見 (必須)	<ol style="list-style-type: none">1 広範囲に分布する紅斑に加え体表面積の 10%を超える水疱・びらんがみられる．外力を加えると表皮が容易に剥離すると思われる部位はこの面積に含める (なお，国際基準に準じて体表面積の 10～30%の表皮剥離は，SJS / TEN オーバーラップと診断してもよい)2 発熱がある．3 以下の疾患を除外できる．<ul style="list-style-type: none">ブドウ球菌性熱傷様皮膚症候群 (SSSS)トキシックショック症候群伝染性膿痂疹急性汎発性発疹性膿疱症 (AGEP)自己免疫性水疱症
副所見	<ol style="list-style-type: none">1 初期病変は広範囲にみられる斑状紅斑で，その特徴は隆起せず，中央が暗紅色の flat atypical targets もしくはびまん性紅斑である．紅斑は顔面，頸部，体幹優位に分布する．2 皮膚粘膜移行部の粘膜病変を伴う．眼病変では偽膜形成と眼表面上皮欠損のどちらかあるいは両方を伴う両眼性の急性結膜炎がみられる．3 全身症状として他覚的に重症感，自覚的には倦怠感を伴う．口腔内の疼痛や咽頭痛のため，種々の程度に摂食障害を伴う．4 病理組織学的に表皮の壊死性変化を認める．完成した病像では表皮の全層性壊死を呈するが，軽度の病変でも少なくとも 200 倍視野で 10 個以上の表皮細胞 (壊) 死を確認することが望ましい．
