

厚生労働科学研究費補助金
政策科学総合研究事業（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

薬局ヒヤリハット事例に対する安全管理対策評価に関するAI研究

研究代表者 岡本 里香
京都大学大学院医学研究科 人間健康科学系専攻 ビッグデータ医科学系分野
特定准教授

研究要旨

独立行政法人医薬品医療機器総合機構（以下、PMDA）では、公益財団法人日本医療機能評価機構（以下、評価機構）が薬局ヒヤリ・ハット事例収集・分析事業（以下、本事業）に基づき収集・分析・公表した「ヒヤリ・ハット事例」に対し、医薬品の名称・包装等の観点から安全対策を講じる必要がないか検討を行っている。本事業は、全国の薬局からの事例を収集・分析し、薬局における医療安全対策に有用な情報を共有するなど、医療安全対策の一層の推進を図ることを目的として行われている。収集される事例は、薬局で発生した調剤や疑義照会等に関するヒヤリ・ハット事例であるが、例えば、調剤に関する事例のうち、薬剤の名称の視覚的、音韻的な類似に起因したことで薬剤取違えた等の事例の場合、PMDAでは製造販売会社に対し、薬剤の取違えを防ぐための注意喚起の必要性等について指導する、といった医薬品の物的要因に対する安全管理対策の評価・検討している。しかしながら、評価機構の薬局ヒヤリ・ハット事例報告数は増加しており、ここ数年では、1年で数10万件が報告されている。この中から対策の必要性を検討しなければならない事例を抽出するだけでも、かなりの労力と言える。

PMDAでは「薬局ヒヤリ・ハット事例」に対して、現在、人による目で、評価を5段階に分類し、安全対策の必要な事例を抽出している。本研究では、この分類を人工知能（以下、AI）が行えるようにすることを目的としている。過去のPMDAにおける評価では、評価機構が公表したの報告に対して、対策を検討する事例は、近年では評価機構が公表した報告の約5.0%程度であり、ほとんどがヒューマンエラーや情報不足の事例であることから、これらを1次スクリーニングとしてAIで分類するだけでも、PMDAにおける労力は軽減され、対策を検討しなければならない事例に注力して安全管理対策を講じることが可能となる。

本研究では評価分類モデルのアルゴリズム検討を行い、PMDAにおいて実用化するまでのAI開発を目指す。

（研究分担者）

中津井 雅彦

山口大学大学院医学系研究科・医学部附属病院AIシステム医学・医療研究教育センター・特命教授

小島 諒介

京都大学大学院医学研究科・人間健康科学系専攻・ビッグデータ医科学分野・講師

A. 研究目的

独立行政法人医薬品医療機器総合機構（以下、PMDA）では、公益財団法人日本医療機能評価機構（以下、評価機構）が薬局ヒヤリ・ハット事例収集・分析事業に基づき収集・分析・公表したヒヤリ・ハット事例に対し、医薬品の名称・包装等の観点から安全対策を講じる必要がないか検討を行っている。例えば、薬剤の名称の視覚的、音韻的な類似に起因したことで薬剤取違えた等の事例の場合、PMDA では製造販売会社に対し、薬剤の取違えを防ぐための注意喚起の必要性等について指導するといった医薬品の物的要因に対する安全管理対策の評価・検討しており、現状、人による目で評価分類し、安全対策が必要な事例を抽出している。評価機構が公表するヒヤリ・ハット事例は年々増加傾向にあり、PMDA においてこれらに対する安全対策要否の検討・評価は負担が大きくなっている。（図 1）

（図 1）

薬局ヒヤリ・ハット報告数と評価対象事例数の推移

PMDA安全使用対策検討会	2013	2014	2015	2016	2017	2018		2019		2020		2021
						前期	後期	前期	後期	前期	後期	前期
評価機構の報告数	6,497	5,387	4,923	4,865	5,151	3,457	20,684 (*1)	34,373 (*2)	92,741 (*3)	97,707 (*4)	23,575	40,186 (*5)
PMDA評価対象	<ul style="list-style-type: none"> * 事例のうち ✓ 規格・剤形間違い ✓ 薬剤取違え ✓ その他及び疑義照会 						評価対象事例の抽出方法の検討のため評価事例なし					
	3,670	3,019	2,959	2,893	3,356	2,424		1,255	2,369	2,416	575	2,959
	調査の結果、対策を講じたもの事例数は非常に少なく、ほとんどが情報不足、ヒューマンエラーの事例である											
対策検討事例数	0	6	12	12	20	11	-	57	122	161	33	148
評価対象事例に占める割合	0.2%	0.2%	0.4%	0.4%	0.6%	0.5%	-	4.5%	5.1%	6.7%	5.7%	5.0%

(*1) 増加背景として、H27年10月公表「患者のための薬局ビジョン」、H29年3月公表「『患者のための薬局ビジョン』実現のためのアクションプラン検討委員会報告書」の影響、さらに、平成30年度診療報酬改定における地域支援体制加算の新設が本事業への参加と報告の増加に影響していると考えられる。

(*2) これまで6か月分の報告を対象としてきたが、H30年度後期のPMDA評価をスキップしたため、9か月間分の報告対象（評価機構報告H30年1～9月分）

(*3) 7か月間分の報告対象（評価機構報告2018年10～2019年4月分）

(*4) 8か月間分の報告対象（評価機構報告2020年5～12月分）

(*5) 評価機構報告 2020年3月17日～2020年7月31日

PMDA における評価は、評価機構が公表する事例のうち、「規格・剤形間違い」「薬剤取違え」「その他及び疑義照会」として報告された事例を抽出し、各事例の内容を確認・評価し、次の評価 1～5 に分類している。

評価 1：医薬品の安全使用に関して製造販売業者等による対策が必要又は可能と考えられた事例

評価 2：製造販売業者等により既に対策がとられているもの、もしくは対策を既に検討中の事例

評価 3：製造販売業者等によるモノの対策は困難と考えられた事例（ヒューマンエラー、ヒューマンファクター）

評価 4：製造販売業者等によるモノの対策は困難と考えられた事例（副作用、情報不足等）

評価 5：その他（処方箋等からの保険者番号等の転記ミスや調剤報酬の算定誤り等）

本研究では、PMDA における評価 1～5 の分類を AI に実施させることにより、PMDA における本業務の負担を軽減し、ヒヤリ・ハット事例の増加に対しても一貫性のある評価を行うことを目的とする。（図 2）

（図 2）

背景

～PMDAの医療安全ヒヤリ・ハット評価の流れ～

10万近い
報告数

公益財団法人 日本医療機能評価機構がHPで公表する「薬局ヒヤリ・ハット事例」

- 事例のうち
- ✓ 規格・剤形間違い
 - ✓ 薬剤取違え
 - ✓ その他及び疑義照会

PMDAにおいて、医薬品の物的要因に対する安全管理対策を評価・検討

AIに評価させ
られないか？

公表内容から以下の内容に分類し、安全対策が必要ないかを検討

- 評価 1. 医薬品の安全使用に関して製造販売業者等による対策が必要又は可能と考えられた事例
評価 2. 製造販売業者等により既に対策がとられているもの、もしくは対策を既に検討中の事例
評価 3. 製造販売業者等によるモノの対策は困難と考えられた事例（ヒューマンエラー、ヒューマンファクター）
評価 4. 製造販売業者等によるモノの対策は困難と考えられた事例（副作用、情報不足等）
評価 5. その他（処方箋等からの保険者番号等の転記ミスや調剤報酬の算定誤り等）

B. 研究方法

初年度（2020年度）の評価分類モデルの精度向上に関する検討の結果を以下にまとめる。

- (1) 特徴量：KEGGを用いた複数名称と薬効の特徴量化により精度向上が認められた。劇薬以外の重篤な健康被害に関する情報を特徴量に含める必要性が示唆された。
- (2) 分類評価スキーム：2step法において、両段階にテキスト情報を特徴量と含めた場合に精度向上に最も寄与し、no-step法よりも有効であった。
- (3) サンプルング方法：Random Under Sampling より Random Over Sampling が有効と考えられた。
- (4) 自然言語処理技術：テキスト情報に対して word2vec よりも Universal Sentence Encoder が有用であった。

2021年度は、これまでに作成した評価分類モデルが新様式データに対して再現できるかを検討した。

- ① 新様式データのPMDA評価結果の分類パターンを検出した。
→評価機構が報告した2020年3月17日から2020年7月31日までの新様式事例の内訳が、様式が変更したことに伴う、評価機構の報告事例からのPMDAによる抽出の結果として、PMDAの評価1から5の事例配分に旧様式に比して不均衡がないかを確認する。
- ② 上記①より、ラベル4及び5の事例が極端に少ないことから、no-step法でも高い分類精度の可能性が考えられたため、no-step法と2step法の比較検討（閾値0.5）を実施した。評価指標は precision, recall, f1-score 及び AUC precision recall curve を用いた。

さらに、精度向上等に向けた検討として、以下の(1)、(2)の学習データについて追加検討を、新様式データに対して no-step 法（4値分類として出力し最も良い指標値を採用）、Random Over Sampling, Universal Sentence Encoder の使用条件（図3）で、学習データ KEGG と T 辞書の比較及び特徴量としてハイリスク薬を含めた場合を検討した。評価指標は precision, recall, f1-score 及び AUC precision recall curve を用いた。

実施した。

- (1) T辞書：商用利用の観点で、KEGGから有限会社ティ辞書企画が販売する医薬品情報データに関する辞書（以下「T辞書」）の変更を検討する。収録語に、医薬品名（一般名、商品名、略語等）、薬効名、疾患名（病名・症状、副作用用語等）、独自で整理された同義語等を含み、同義語についてはKEGGにない情報。
- (2) ハイリスク薬情報：劇薬以外の重篤な健康被害に関する情報として、PMDAでは、厚生労働省保険局による「特定薬剤管理指導加算等の算定対象となる薬剤一覧」を活用しているため、当該薬剤一覧にある薬剤を特徴量に追加検討する。

<図 3>新様式データに対する特徴量等の条件

特徴量	これまでの評価分類モデル	今年度提案
販売名のみ類似性	編集距離など複数の手法を利用	編集距離など複数の手法を利用
一般名まで含めた類似性	KEGGを利用	T辞書を利用して特徴量化
T辞書の同義語を利用した類似性	-	T辞書を利用して特徴量化
テキスト情報 (Universal Sentence Encoder)	事例の詳細 背景要因 改善策 発生意因 をencodeして連結	事例の詳細 背景要因 改善策 発生意因 をencodeして連結
KEGGの薬効情報	KEGGを利用	T辞書を利用して特徴量化
ハイリスク薬情報	-	「特定薬剤管理指導加算等の算定対象となる薬剤一覧」を利用して特徴量化

(倫理面への配慮)

本研究で使用するデータは個人情報を含まず、公表済のものを使用しているため、倫理面での配慮は特にない。

C. 研究結果

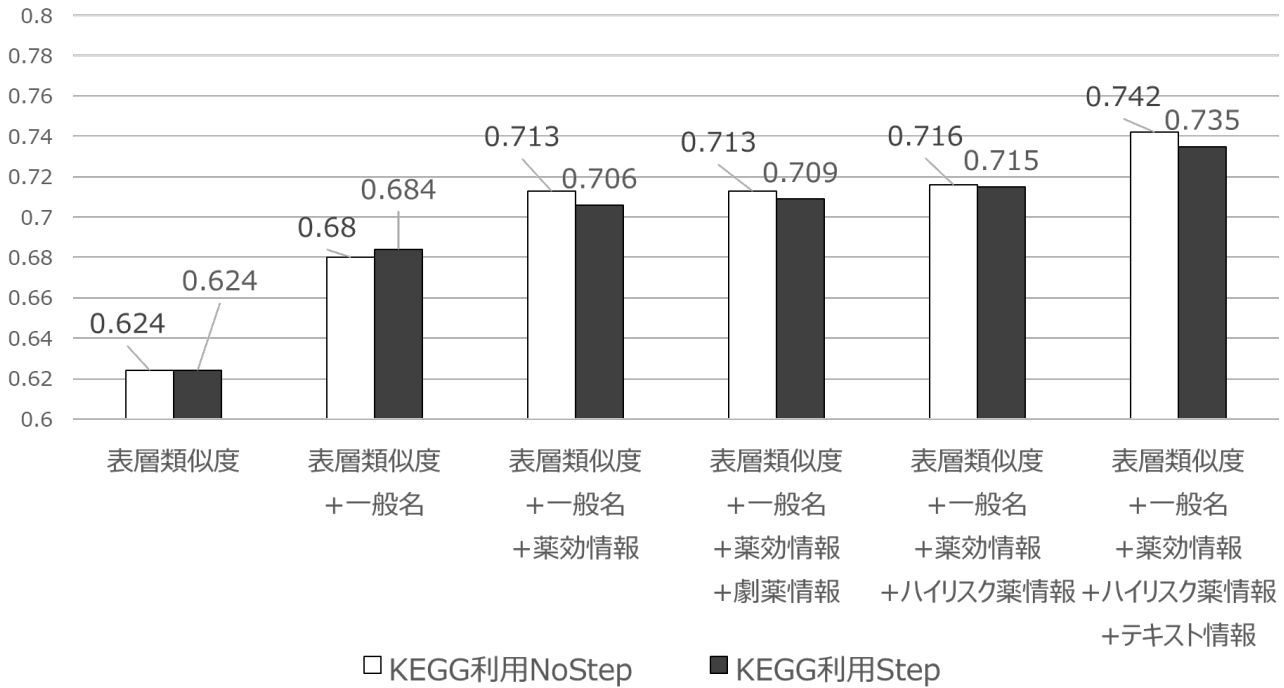
新様式データに対して、no-step 法と 2step 法の比較検討 (閾値 0.5) を実施した結果を、ハイリスク薬情報の特徴量追加検討も含めて示す (図 4)。

2step 法による recall, f1-score の精度向上がラベル 3 に認められた。しかし、この結果は、あくまで閾値 0.5 での結果であり、AUC の結果 (図 5) では、no-step 法の場合に精度向上が認められ、全体の性能としては no-step 法で表層類似度+一般名+ハイリスク薬+薬効情報+テキスト情報の特徴量の組合せの場合に AUC は最も高く、実用化に向けては、実務上許容できる recall を参考に no-step 法か 2step 法かを選択する必要があると考える。

<図 4>新様式データに対する no-step 法と 2step 法の比較 (各評価指標)

feature	label	No-Step			2Step		
		precision	recall	f1score	precision	recall	f1score
表層類似度	1,2	0.62	0.64	0.63	0.63	0.64	0.63
	3	0.98	0.80	0.88	0.97	0.98	0.97
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名	1,2	0.97	0.64	0.77	0.97	0.64	0.77
	3	0.98	0.82	0.89	0.97	0.99	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報	1,2	0.99	0.66	0.79	0.99	0.65	0.78
	3	0.98	0.82	0.89	0.97	1.00	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報 +劇薬情報	1,2	0.99	0.65	0.78	0.98	0.66	0.79
	3	0.98	0.82	0.89	0.97	1.00	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報 +ハイリスク薬情報	1,2	0.98	0.66	0.79	0.99	0.65	0.78
	3	0.98	0.82	0.89	0.97	1.00	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報 +ハイリスク薬情報+テキスト情報	1,2	0.96	0.64	0.77	0.96	0.63	0.76
	3	0.97	0.99	0.98	0.97	0.99	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.13	0.15	0.14	0.00	0.00	0.00

<図 5>新様式データに対する no-step 法と 2step 法の比較 (AUC PrecisionRecallCurve)

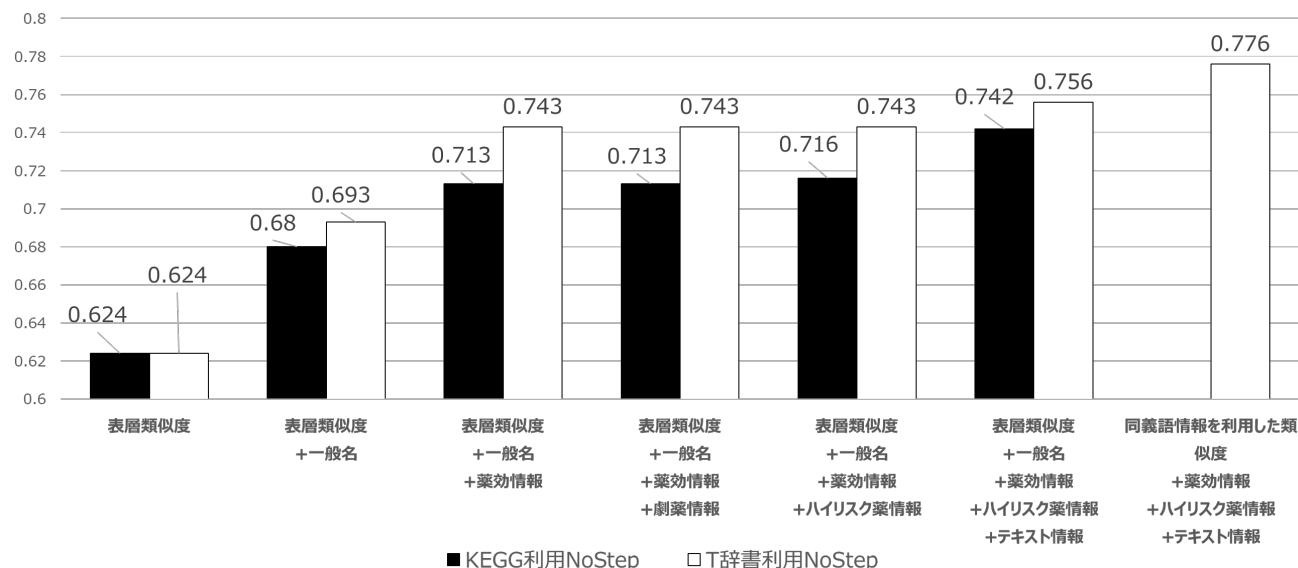


次に、上記の結果で最適と考えられた分類スキームの no-step 法、そして、これまでに最適であることを検討したサンプリング方法として Random Over Sampling 法を組合せ条件として、KEGG と T 辞書の比較を上述の図 3 の内容で行った。その結果、precision, recall, f1-score では既存手法と T 辞書利用で差はほとんどないことから、T 辞書利用で分類精度の低下はないことが示され (図 6)、AUC では、T 辞書利用は既存手法よりも高く、T 辞書に含まれる同義語を特徴量として含めた場合に、最も高い AUC が得られた (図 7)。

<図 6>新様式に対する KEGG (既存手法) と T 辞書との比較 (各評価指標)

feature	label	既存手法			T辞書利用		
		precision	recall	f1score	precision	recall	f1score
表層類似度	1,2	0.62	0.64	0.63	0.62	0.64	0.63
	3	0.98	0.80	0.88	0.98	0.80	0.88
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.02	0.77	0.04
表層類似度+一般名	1,2	0.97	0.64	0.77	0.96	0.66	0.78
	3	0.98	0.82	0.89	0.98	0.82	0.89
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.02	0.77	0.04	0.02	0.77	0.04
表層類似度+一般名+薬効情報	1,2	0.99	0.66	0.79	0.98	0.68	0.80
	3	0.98	0.82	0.89	0.98	0.82	0.89
	4	0.00	0.00	0.00	0.01	0.60	0.02
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報+劇薬情報	1,2	0.99	0.65	0.78	0.99	0.68	0.80
	3	0.98	0.82	0.89	0.98	0.82	0.89
	4	0.00	0.00	0.00	0.01	0.60	0.02
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報+ハイリスク薬情報	1,2	0.98	0.66	0.79	0.98	0.68	0.80
	3	0.98	0.82	0.89	0.98	0.82	0.89
	4	0.00	0.00	0.00	0.01	0.60	0.02
	5	0.02	0.77	0.04	0.00	0.00	0.00
表層類似度+一般名+薬効情報+ハイリスク薬情報+テキスト情報	1,2	0.96	0.64	0.77	0.95	0.61	0.75
	3	0.97	0.99	0.98	0.97	0.99	0.98
	4	0.00	0.00	0.00	0.00	0.00	0.00
	5	0.13	0.15	0.14	0.00	0.00	0.00
T辞書同義語+薬効情報+ハイリスク薬情報+テキスト情報	1,2				0.98	0.67	0.80
	3				0.98	1.00	0.99
	4				0.00	0.00	0.00
	5				0.13	0.08	0.10

<図 7>新様式に対する KEGG と T 辞書との比較 (AUC PrecisionRecallCurve)



また、KEGG と T 辞書でそれぞれ最も分類精度が良い特徴量組合せの分類事例数を比較したところ、KEGG と、T 辞書利用で正しく分類されたラベル 1・2 の事例数は、KEGG では 95 事例に対し、T 辞書では 99 事例であり、T 辞書を利用することによりラベル 1,2 への分類精度が高まった (図 8)。KEGG でラベル 1,2 と分類されるべき事例はラベル 3 に誤分類されていることから、この 4 事例を精査したところ、KEGG には含まれていない情報として、T 辞書に含まれている同義語が、名称変更前後の販売名に関する取違え事例の分類に寄与していることが分かった。

<図 8> 最も分類精度が良い特徴量組合せの KEGG と T 辞書での分類事例数

NoStep法でKEGGによる特徴量化
表層類似度+一般名
+ハイリスク薬情報+薬効情報+テキスト情報

NoStep法でT辞書による特徴量化
同義語情報利用の類似度
+ハイリスク薬情報+薬効情報+テキスト情報

		予測			
		1,2	3	4	5
真	1,2	95	53	0	0
	3	4	2767	3	14
	4	0	10	0	0
	5	0	9	2	2

		予測			
		1,2	3	4	5
真	1,2	99	49	0	0
	3	2	2776	3	7
	4	0	10	0	0
	5	0	12	0	1

※予測結果の確率の内、最も高いラベルを採用

<図 9>T 辞書利用で正しく分類されたラベル 2 の事例

薬1	薬2	考察
アレロック	シルニジピン	シルニジピンの旧称がアレロックであり、T辞書に同義語として登録されていた
シルニジピン	オロパタジン	シルニジピンの旧称がアレロックであり、オロパタジンの同義語にアレロックが登録されていた
セルシン	プロマゼパム	プロマゼパムの旧称がセニランであり、T辞書に同義語として登録されていた
イソコナールR	一硝酸イソソルビド	2020年6月にイソコナールRは名称変更で硝酸イソソルビドとなっているが、クローラしたKEGGに未反映だった

しかし、T 辞書をした場合でも、図 8 のとおり、ラベル 1,2 をラベル 3 に誤分類する事例が 49 事例ある。49 事例の内容を図 10 のようにケース分けを行ったところ、ケース 1 では、情報提供文書の発出により、過去にラベル 3 として判断されていたものでも、その後取り違いに関する情報提供文書が発出された場合、ある時点から評価がラベル 3 からラベル 1,2 に変更になったケースであり、AI ではカバーできない要因によ

る誤分類が存在することが分かった。

<図 10>

ケース No.	ケース内容	事例数
1	情報提供文書発出により評価分類が変更になっているケース	30
2	AI 視点で評価 2・3 の判断を悩むケース	5
3	報告内の薬剤名が誤記載のケース	2
4	新規薬剤組合せのケース	1
5	過去 PMDA 評価結果について確認が必要なケース	11

D. 考察

旧様式データでは、PMDA 評価 1～5 で大半が評価 4 及び 5 に該当する事例が多い構成であったが、新様式データでは、様式変更に伴い、結果的に PMDA 評価 4 及び 5 がほとんど含まれていないことがわかった。そのため、これまでの検討で評価 1～5 から評価 1 及び 2 を抽出するための対処として no-step 法と 2-step 法の検討を行ってきたが、新様式データに対しては、no-step 法か 2-step 法のいずれも、分類スキームとして実施している内容が変わりがほぼなくなり、評価結果についてもほぼ差が認められず、むしろ no-step 法で AUC が向上する結果が得られた。実用化に向けては、実務上許容できる recall を参考に no-step 法か 2step 法かを選択する必要があるが、これまでに検討した評価分類モデルが新様式データにも適用可能であることが示された。

またハイリスク薬の特徴量として、これまで劇薬情報に注目していたが、「特定薬剤管理指導加算等の算定対象となる薬剤一覧」の薬剤を特徴化することにより、分類精度が向上することが確認できた。更に、T 辞書の利用は KEGG の代用として問題なく、特に T 辞書に含まれる同義語は、これまでの学習データで欠如していた「名称変更前後の販売名」に寄与することから、KEGG よりも有用であることが示された。

一方で、AI ではカバーできない要因による誤分類が存在することから、例えば、実運用の際は機械学習による手法のみでなく、ラベル 1,2 とする取り違いペアを事前に定義し一部はルールベースによる判断を組み入れるなどの運用上の工夫が必要であると考えられた。

E. 結論

昨年度までに作成し、最適な条件を検討してきた評価分類モデルについて、新様式データに対しても適用可能であることが示された。旧様式では、報告様式上、フリーテキストで記入する箇所があったことから、評価分類モデルの作成過程で、テキスト情報に対する特徴量化、分類スキーム、サンプリング方法などの検討を行ってきたが、新様式の報告内容のチェックボックス化に伴う簡略化されたデータに対しても、評価分類モデルはこれまでの精度を落とすことなく、分類することができた。むしろ、旧様式データに対して、分類の取りこぼしを避けるための手法として採択した分類スキームの 2-step 法よりも、より簡易な no-step 法が新様式データに対して効果があったことは、よりシンプルなアルゴリズムを使用できる点で、医療安全評価の AI を使用する側での安心感に繋がると考える。また、PMDA での実装化を考慮して、KEGG から T 辞書への利用変更を検討した結果は、評価分類モデルの精度向上に繋がる結果が得られたことは非常に興味深い。一方で、AI ではカバーできない要因による誤分類が存在することから、PMDA での運用では、1 次スクリーニングという位置づけで評価分類モデルを運用することに加え、特徴量の追加・調整や閾値の調整などで AI 機能に対する精度向上も目指しつつ、AI ではカバーできない誤分類に対しルールベースの分類を組み合わせるなど、最終年度（2022 年度）において、実運用に向けたアルゴリズム全体の検討を進める必要があると考える。

本研究の検証結果は、テキスト情報を含む評価機構データ以外の医療情報に対しても活用できる可能性があり、医療情報のリスク抽出・分析等の研究開発にも重要な成果といえる。また、T 辞書といった医療関連辞書の活用により精度向上が認められ、本研究成果を feedback することにより医療関連辞書の性能向上にも寄与できる可能性があると考えられる。

F. 健康危険情報 なし

G. 研究発表

- 岡本里香, 中津井雅彦, 小島諒介 薬局ヒヤリ・ハット事例に対する安全管理対策評価に関する AI 開発 第 7 回日本医療安全学会学術総会 (オンライン) 2021 年 5 月 25 日

2. 岡本里香, 中津井雅彦, 小島諒介 薬局ヒヤリ・ハット事例に対する安全管理対策評価に関する AI 開発 第 11 回レギュラトリーサイエンス学会学術大会 2021 年 9 月 17 日

H. 知的財産権の出願・登録状況 なし