

自然言語処理を用いた医療インシデント報告の患者影響度推定

栗井実緒†

藤田 桂英†

Zoie Shui-Yee Wong‡

† 東京農工大学

‡ 聖路加国際大学 公衆衛生大学院

1 はじめに

近年、自然言語処理技術の医療分野への応用に期待が高まっている。患者安全の分野においても例外ではなく、医療上の意図しないアクシデントを記録した医療インシデント報告のテキストデータを利活用することで、より安全な医療へと繋げるための様々な試みが行われている [1][2].

本研究では、聖路加国際病院において収集されたインシデント報告の患者影響度を、自動で推定する分類器を提案する。

2 医療インシデント報告データセット

2.1 データセット 1 - 聖路加データ

聖路加国際病院において、2009年4月から2018年3月に収集された、「薬剤関連」に分類されるインシデント報告 6628 件 (いずれも匿名化されたもの) を用いる。本研究で分類を行う「患者影響レベル」は、インシデントによる患者への影響の深刻度を7段階 (○, I, II, IIIa, IIIb, IV, V) で示したものである。クラスごとのデータ数を表 1 に示す。極端なクラス間の偏りを考慮し、実験では表 1 中の「データ数 (削減後)」の項目に示した数のデータのみ (合計 1941 件) を使用し、患者影響度 ○, I, II, IIIa の 4 クラス分類タスクとして扱う。

2.2 データセット 2 - JQ データ

公益財団法人日本医療機能評価機構 (JQ) が運営している医療事故情報収集等事業によって収集された、46503 件の医療事医療機関事故情報及びヒヤリ・ハット事例情報データである [3].

3 医療インシデント報告の自由記述欄における影響度の自動分類

本研究では、機械学習によって患者影響度の4クラス分類 (○, I, II, IIIa) タスクに試みる。インシデント報告に記載されているテキストデータから、(1) 単語文書行列 (2) トピックモデルの二種類の方法で特徴量を生成し、分類器への入力とする。分類器は線形 SVM (liner), 非線形 SVM (RBF), ランダムフォレストを比較する。いずれもグリッドサーチによって最適なパラメーターを

表 1: 聖路加データ - 患者影響レベルの分布

患者影響レベル	データ数 (オリジナル)	データ数 (削減後)
○	520	520
I	5081	500
II	389	389
IIIa	532	532
IIIb	29	0
IV	0	0
V	0	0
なし	77	0

表 2: 実験 1 - 聖路加データのクラス分類

特徴量	分類器	精度
Bag of Words	Random Forest	0.6842
TF-IDF	Random Forest	0.6878
Binary	SVM(RBF)	0.6924

探索し、5 分割交差検証によって分類器の性能を評価する。

実験 1: 単語文書行列による特徴量生成

Bag of Words, TF-IDF, Binary (Bag of Words のように単語の出現回数を考慮せず、「単語が存在するか否か」のみによって、1 と 0 で表現する方法) の 3 種類の方法によって聖路加データのテキストデータをベクトル化したものを、特徴量とし、分類器への入力とする。

実験の結果、特徴量ごとに得られた最高精度と、その場合に利用した分類器を表 2 に示す。Binary を特徴量とし、SVM(RBF) を分類器とした場合に、最も高い精度 ($accuracy = 0.6924$) が得られた。各クラスの F 値はレベル ○ が 0.7926, レベル I が 0.6393, レベル II が 0.4846, レベル IIIa が 0.7977 が得られた。

実験 2: トピックモデルによる特徴量生成

トピックモデル (LDA) を用いて、聖路加データの特徴量を生成する。トピックモデルに JQ データのテキストデータを学習させる。学習後のトピックモデルによって、聖路加データのテキストデータをトピックにもとづいた特徴量に変換し、分類器への入力とする。トピック数 $K = 10, 50, 100, 300, 500, 800, 1000$ を比較する。

表 3: 実験 2 - 聖路加データのクラス分類

分類器	トピック数 K	精度
SVM	500,800	0.6682
Random Forest	300	0.6801

表 4: 実験 3 - 聖路加データのクラス分類

追加した特徴量	精度	t 検定
構造化データなし	0.6924	-
経験年数	0.6580	0.0120
部署	0.6899	0.2276
時間帯	0.6910	0.1870
職種	0.6863	0.0757
構造化データを全て追加	0.6641	0.0131

実験の結果、分類器ごとに得られた最高精度と、その際のトピック数 K を表 3 に示す。トピック数 $K = 300$, $t = 80\%$ として、分類器としてランダムフォレストを用いた場合に最も高い精度 ($accuracy = 0.6801$) が得られた。また、各クラスの F 値はレベル O が 0.7471, レベル I が 0.6423, レベル II が 0.4054, レベル IIIa が 0.8063 が得られた。

実験 3, 実験 4 では、実験 1 と実験 2 を通して最高精度が得られた手法 (単語文書行列 (Binary) を特徴量とし, SVM(RBF) を分類器とした場合 *) を適用する。

実験 3: 構造化データを特徴量に追加した分類

前述の手法 (*) に、テキストデータ以外の情報 (構造化データ) を特徴量として追加し、分類精度への影響を解析する。インシデント報告に記載されている「経験年数」「部署」「時間帯」「職種」の項目を特徴量に追加する。5 分割交差検証の結果を実験 1 と比較して t 検定を行うことで、精度の差が統計的に有意か検証を行う (有意水準 $p < 0.05$)。

表 4 は聖路加データのクラス分類の結果を示している。表 4 から、構造化データなしが最も高い精度を示しており、いずれの構造化データを特徴量に追加しても精度の向上はなかった。

実験 4: 外部データへの手法の適用

実験 1~実験 3 の手法を他のインシデント報告データセットに適用し、データセットに依存しない適用可能性を検証する。前述の手法 (*) と、実験 3 と同様の特徴量を JQ データの分類に適用する。本実験では JQ データにおける「影響度」の項目を推定し、推定する影響度は「軽微な処置・治療が必要もしくは処置・治療が不要と考えられる (クラス 0)」731 件、「濃厚な処

表 5: 実験 4 - JQ データのクラス分類

追加した特徴量	精度	t 検定
構造化データなし	0.5116	-
発生時間帯	0.5148	0.0807
事故に直接関連する疾患名 ★	0.5363	0.0116
事故の内容 ★	0.5463	0.0073
患者の年齢	0.5157	0.0979
患者の性別 ★	0.5199	0.0200
患者の数 ★	0.5148	0.0022
患者区分	0.5166	0.1068
発生場面 ★	0.5454	0.0069
発生要因 (その他)	0.5121	0.4582
発生要因 (ヒューマンファクター)	0.5244	0.0990
発生要因 (当事者の行動に関わる要因)	0.5157	0.1334
発生要因 (環境・設備機器)	0.5180	0.1120
発見者	0.5162	0.1704
直前の患者の状態 ★	0.5390	0.0242
種類	0.5148	0.3292
構造化データを全て追加 ★	0.5832	0.0002

置・治療が必要であると考えられる (クラス 1)」628 件、「死亡もしくは重篤な状況に至ったと考えられる (クラス 2)」834 件の 3 クラス (合計 2193 件) で構成される。

表 5 に聖路加データで訓練した学習器を JQ データに適用させた場合の精度を示す。前述の手法 (*) を適用した場合の精度は 0.5116 であった。聖路加データの場合よりも分類精度が低い理由として、JQ データは比較的記述量が少ないため、テキストデータ含まれる情報量に差があることが考えられる。構造化データを特徴量に追加した分類では、表 5 中の ★ の項目を追加した場合に、有意な精度の向上が認められた。構造化データを全て追加した場合に最も高い精度が得られた。

4 まとめ

本論文では、機械学習によるクラス分類によって、医療インシデント報告のテキストデータから、その患者影響度を推定する手法を提案した。また、複数のデータセットにおける評価実験を行い、作成した分類器および分類に有効な項目の検証を行った。

参考文献

- [1] Ying Wang and Enrico et al. Coiera. Using multiclass classification to automate the identification of patient safety incident reports by type and severity. *BMC Medical Informatics and Decision Making*, 17(1):84, 2017.
- [2] Yuko Shiima and Zoie Wong. Classification scheme for incident reports of medication errors. *Studies in health technology and informatics*, 265:113–118, 08 2019.
- [3] 後 信. 我が国の医療安全対策の歩みと医療事故、ヒヤリ・ハットの収集事業. In 日本医療機能評価機構 NEWS LETTER 2012 No.4. 公益財団法人日本医療機能評価機構, 2012.