

医薬品有害事象予測における医薬品数の拡張と Voting による予測精度向上の試み

榎阪 優希^{*1} 南崎 和哉^{*2}, 松野 純男^{*2}, 金子 周司^{*3}, 大星 直樹^{*4},
近畿大学 ^{*1}理工学部情報学科, ^{*2}近畿大学 薬学部医療薬学科
^{*3}京都大学 大学院薬学研究科, ^{*4}近畿大学 大学院総合理工学研究科

Expanding the Number of Drugs in the Prediction of Adverse Drug Events and Attempt to Improve Prediction Accuracy by Voting

Yuki Esaka^{*1}, Kazuya Nanzaki^{*2}, Sumio Matzno^{*2}, Shuji Kaneko^{*3}, Naoki
Ohboshi^{*4}

^{*1} Dept of Informatics, Faculty of Science and Engineering, Kindai University

^{*2} Dept of Medical Pharmacy, Faculty of Pharmacy, Kindai University

^{*3} Kyoto University, Graduate School of Pharmaceutical Sciences

^{*4} Kindai University, Graduate School of Science and Engineering

抄録: 独立行政法人医薬品機器総合機構(PMDA)が公開している医薬品副作用データベース(JADER)を活用し、有害事象の予測を行った。650,605 件数の JADER データを、同成分で異なる名称で登録されている医薬品名に対して、ライフサイエンス辞書を用いて医薬品名の統一処理を行なった。その後、JADER データ中の投与医薬品5剤服用時と6剤服用時事例に範囲を限定し、学習・予測を行った。ここでは Support Vector Machine(SVM)、Deep Neural Network(DNN)、Random Forest(RF)の3手法に加えて、アンサンブル学習の一種である Voting を使用し、予測精度の向上を目指した。

各手法のそれぞれの予測結果から、Voting において 0.01 程度の精度向上がみられたことを確認した。

キーワード JADER、Support Vector Machine、Deep Neural Network、Random Forest、Voting

1. はじめに

医療においては患者に複数の医薬品を同時に処方し、治療することがある。しかし、摂取した患者の体質、健康状態、医薬品の相互反応により予想外の有害事象が発現する可能性がある。

我々は先行研究[1]として、JADER[2]に対して投与医薬品数を5剤服用時にデータを絞って学習を行った。Support Vector Machine (SVM)[3]、Deep Neural Network (DNN)[4]、Random Forest (RF)[5]の3手法間での差異の傾向、最も適した学習手法を検討した。その結果、各手法にあまり性能差がなく、3手法とも精度が8割程度であった。

本稿では、2004年から2020年10月までのデータを活用し、投与医薬品数を5剤と6剤服用時のケースに適応範囲を拡大、複数の予測手法を組み合わせることで予測精度の向上を目指し、副作用の予測を行った。今回選択した機械学習の手法は SVM、DNN、RF に、アンサンブル学習の

一種である Voting[5,6]を加えた4手法である。

2. 方法

1) 前処理・データクリーニング

医薬品名の表記を統一するため、ライフサイエンス辞書[7]を用いて、医薬品名の統一を行なった。ライフサイエンス辞書にある医薬品名と一致、または類似度が高いものに関しては Web スクレイピングの手法を用いて検索、結果の取得を自動で行い、類似度が低いものに関しては手動で用語の統一を行なった。その結果、9,294種類あった医薬品を4,458種類に統合できた。全データから投与医薬品数5,6剤のデータを抽出、各有害事象のデータ件数に差があることを考慮し、アップサンプリングを行った。

2) 機械学習

対象データについて、RF、SVM、DNN、Voting の4手法を用いて分類を行った。各手法のパラメータを表1に示す。DNN については入

力層、中間層2層、出力層の4層で構成した。なお、パラメータの調節にはグリッドサーチを使用した。

表1 各手法のパラメータ

手法	パラメータ
RF	木の深さ:100 木の数の上限:200 分割指標:エントロピー
SVM	正規化係数:300 カーネル:ガウスカネル カーネルの係数:0.01
DNN	最適化アルゴリズム:アダム L2 正規化係数:0.0001
Voting	分類手法:hard 各手法の重さ:RF:3、SVM:2、DNN:2

3) 評価方法

データ分割数を6として Cross Validation で評価する。各手法の単体でのそれぞれの予測精度と Voting の手法を組み合わせた予測精度を比較し、本研究の評価を行った。各手法の全体における精度、適合率、再現率、F 値の4つの指標について比較評価した。

3. 結果

各手法の予測結果を表2に示す。2020 年 10 月の5剤服用時のデータで RF、SVM、DNN、Voting の4手法を比較した結果は、最大 0.80 で RF と Voting が同じ精度であった。2020 年 10 月の6剤のデータで RF、SVM、DNN、Voting の4手法を比較すると、最大 0.82 で Voting が最も高い精度であった。

表2 各手法の予測結果

		精度	適合率	再現率	F 値
5 剤	RF	0.80	0.81	0.82	0.81
	SVM	0.80	0.81	0.82	0.81
	DNN	0.78	0.79	0.80	0.79
	Voting	0.80	0.81	0.82	0.81
6 剤	RF	0.81	0.82	0.83	0.82
	SVM	0.81	0.82	0.83	0.81
	DNN	0.79	0.80	0.80	0.79
	Voting	0.82	0.82	0.83	0.82

4. 考察

Voting による若干の予測精度の向上は見受けられたが、最大 0.01 に過ぎず、期待したほど

の予測精度の向上結果は得られなかった。原因として、Voting で使用する分類手法の種類や数が不足していることが考えられる。

5. 結語

本研究では、医薬品を服用した時の有害事象について、学習・分類を行い、RF、SVM、DNN、Voting の4手法の比較を行った。その結果、Voting は他の3手法よりも最大で 0.01 ほど高い予測精度であった。これからの課題として Voting で使用する分類器の種類、数を増やせばさらに精度向上が図れるのではないかと考えている。

参 考 文 献

- [1] 筒井凌, 小野田涼, 松野純男, 大星直樹: 医薬品副作用データベースを用いた医薬品相互作用での有害事象推定へ適用する機械学習手法の比較, 医薬品情報学 vol.22(3) pp.125-130,2020
- [2] 医薬品医療機器総合機構 Pharmaceuticals and Medical Devices Agency (PMDA)、 「データセットのダウンロードページ」で提供する情報について、 <https://www.pmda.go.jp/safety/info-services/drugs/adr-info/suspected-adr/0004.html> (閲覧日:2021 年 1 月 31 日)
- [3] 竹内一郎, 鳥山昌幸:サポートベクトルマシン, 講談社, 2015
- [4] 岡谷貴之:深層学習, 講談社, 2015
- [5] Zhi-Hua Zhou:アンサンブル法による機械学習ー基礎とアルゴリズムー, 近代科学社, 2017
- [6] A.Narasimhamurthy : Theoretical Bounds of Majority Voting Performance for a Binary Classification, IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.27(12), 2005
- [7] ライフサイエンス辞書, LIFE SCIENCE DICTIONARY, <https://lsd-project.jp/> (閲覧日:2021 年 1 月 31 日)