

令和2年度

厚生労働科学研究費補助金(化学物質リスク研究事業)

**バイタルサインの統合的評価をエンドポイントとした新規急性経口投与毒性試験方法の開発
-統計学による半数致死量から診断学による概略の致死量への転換-(19KD1002)**

分担研究報告書

分担研究課題 バイタルサインの統合的解析方法(ソフトウェア)の開発

研究分担者 相崎 健一

国立医薬品食品衛生研究所 安全性生物試験研究センター 毒性部 第一室長

研究要旨

バイタルサイン(VS)の統合的評価をエンドポイントとした新規急性経口投与毒性試験方法の開発を目的とした本研究班において、最終的には、計測したVSの諸項目から「診断学」を基盤にした「概略の致死量」を推定するためのAcute Toxicity Vital Signs Score(ATVSS 仮称)を定義し、これを実装したソフトウェアの開発を目的とし、研究を行った。開発中のバイタルサイン(VS)取得デバイスはまだ充分量のデータを生成していないため、完成時に得られるデータと同質のものとして、ヒトの心電図データ(MIT-BIH Arrhythmia Database (mitdb) (<https://physionet.org/content/mitdb/1.0.0/>)より入手)を用いた。汎用性の高い異常検出アルゴリズムとして、人工知能、とりわけ近年、開発研究が飛躍的に進み、代表的な手法となっている深層学習(Deep Learning)に着目し、その判別性能等を検討した。総合的な正答率は93%と期待より低いものであったが、これは異常波(心室異所性拍動)に対するスコアが低いためであり、学習に用いたデータの極端な不均衡(心室異所性拍動レコードが全数の一割しかない)が原因と考えられた。対策としては、学習に用いるデータの選定において正常レコード数を削減して異常レコード数とのバランスをとるといった単純な手法で対応可能であり、さらに必要に応じて他の学習モデル、例えば異常検知に有効とされている自己符号化器(autoencoder)や深層距離学習(deep metric learning)などの導入と組み合わせを行うことで、性能向上が可能と考えられる。来年度は、従来からの統計を含む数理的手法や機械学習技術を導入して、ATVSS定義のために必要な、多項目・複数クラス分類の入力データに対応した異常検知モデルの構築を進め、完成予定のVS取得デバイスによる実データを得て、性能評価を試みる計画である。

A. 研究目的

バイタルサイン(VS)の統合的評価をエンドポイントとした新規急性経口投与毒性試験方法の開発に際して、計測した VS の諸項目から「診断学」を基盤にした「概略の致死量」を推定するための Acute Toxicity Vital Signs Score(仮称)を定義し、これを実装したソフトウェアの開発を目的とする。

B. 研究方法

B-1 学習・評価用データ

開発中のバイタルサイン(VS)取得デバイスはまだ充分量のデータを生成していないため、完成時に得られるデータと同質のものとして、ヒトの心電図データ(MIT-BIH Arrhythmia Database (mitdb) (<https://physionet.org/content/mitdb/1.0.0/>)より入手)を用いた。

B-2 解析計算及びソフトウェア生成

異常検出に利用し得る人工知能アルゴリズムのコーディングについては、関連ライブラリが充実している Python 言語(ver.3.6.9)を使用した。機械学習ライブラリとしては Chainer (ver.7.7.0)、GPU 処理ライブラリとして CUDA (ver.10.2.89)、mitdb からの心電図データのダウンロード・処理ライブラリとして WFDB (ver.3.1.1)、その他、Numpy (ver.1.19.4)、Matplotlib (ver.3.3.3)を使用した。Python スクリプト実行環境としては Jupyter Notebook (ver.6.1.5)を使用した。

B-3 計算精度検証:

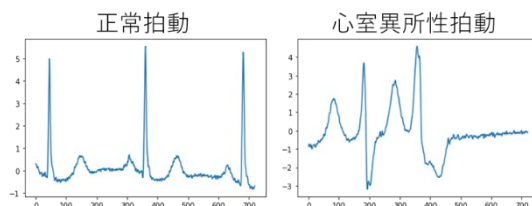
精度検証は必要に応じて Excel(USA Microsoft Corporation)や R 言語(オープンソース R Development Core Team)で実施し、浮動小数点誤差以上の乖離がないことを確認した。

C. 研究結果

現在開発中のバイタルサイン(VS)取得デバイスが完成した際に生成されるデータは、主に一次元の波形データになると予想されるため、今年度はヒトの心電図データをサンプルとして用い、異常検出アルゴリズムの評価や、最終目標である「概略の致死量」を推定するための Acute Toxicity Vital Signs Score(仮称)の「定義」を行う方策を検討した。

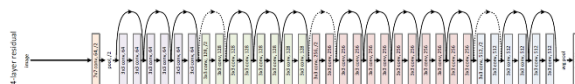
一般的に新規開発デバイスによる計測値処理では、実データを得てからでないと測定値分布も推定するしかなく、数理的な手法では閾値など具体的な基準値を設定するのは困難である。このため、汎用性の高い異常検出アルゴリズムとして、人工知能、とりわけ近年、開発研究が飛躍的に進み、代表的な手法となっている深層学習(Deep Learning)に着目し、その判別性能等を検討した。

まず学習及び評価用データとして、MIT-BIH Arrhythmia Database (mitdb)に収録されていた 47 人の心電図データ、延べ 24 時間のシグナルデータを入手した。このデータベースでは予め各 R 波ピークに AAMI (Association for the Advancement of Medical Instrumentation) 推奨基準によるラベルが付与されており、これを元に個々のレコードに分割し、正常拍動若しくは心室異所性拍動にラベリングした。



前処理として、ペースメーカー等によるノイズが混入したレコードや重複レコードを取り除き、約 9 万レコードのサンプルデータを得た。なお、これらの 9 割が正常拍動、残り 1 割が心室異所性拍動であった。

深層学習で用いるネットワーク構造としては、画像認識タスクで著名な畳み込みニューラルネットワーク(CNN: Convolutional Neural Networks)ベースの ResNet34 (arXiv:1512.03385)を採用した。



ただし、画像認識時の二次元畳み込みではなく、入力データである心電図に合わせて一次元畳み込みに設定し、上記 9 万件の拍データのうち半数を用いて教師あり深層学習を実施した。

学習に用いなかった残り半数の心拍データに対して、学習済み CNN で評価を実行した結果、正答率 (accuracy) は 93.0% であった。内訳としては正常拍動レコードに対する予測スコアは適合率 (precision) 0.99 / 再現率 (recall) 0.93 であったが、心室異所性拍動レコードに対しては同 0.50 / 0.89 であった。

accuracy		0.930	
	precision	recall	f1 score
正常拍動	0.99	0.93	0.96
異常拍動	0.50	0.89	0.64

D. 考察

バイタルサインの統合的評価のためのソフトウェア開発において、今年度は代表的な畳み込みニューラルネットワーク (CNN) モデルを採用し、一般的な教師あり深層学習による一次元の時系列データ (評価用サンプルとして心電図データを利用) の判別性能評価を実施した。総合的な正答率は 93% と期待より低いものであったが、これは異常波 (心室異所性拍動) に対するスコアが低いためである。これは学習に用いたデータの極端な不均衡 (心室異所性拍動レコードが全数の一割しかない) が原因と考えられる。対策としては、学習に用いるデータの選定において正常レコード数を削減して異常レコード数とのバランスをとる (アンダーサンプリング) といった単純な手法で対応可能である。さらに必要に応じて他の学習モデル、例えば異常検知に有効とされている自己符号化器 (autoencoder) や深層距離学習 (deep metric learning) などの導入と組み合わせを行うことで、性

能向上が可能と考えられる。今回の検討は一種類のデータ、且つ正常 / 異常の単純分類であったが、機械学習技術は本来、多種類のデータ入力と複数クラスへの分類 (出力) に対応する柔軟なものであり、本研究の目標である、計測したバイタルサイン (VS) の諸項目 (= 多種類データ) から「診断学」による分類 (= 複数クラス分類) に基づく統合評価の実現に必要な要件を充足する技術として、大変有望な手法と考えられる。なお機械学習技術を判別に利用する場合の最大の懸念は、特徴抽出工程をヒトが行わず機械が自動的に処理してしまうために、判断根拠がブラックボックス化することが想定される。ブラックボックスを内包する毒性試験法では、その判定結果への信頼度が低下するだけでなく、VS 変動に関与する毒性機序研究においても支障を来す。幸い、人工知能研究分野において機械学習の判断根拠を説明する技術が急速に開発されつつあり、将来的にはブラックボックス問題は解消される見込みである。来年度は、従来からの統計を含む数理的手法や機械学習技術を導入して、ATVSS 定義のために必要な、多項目・複数クラス分類の入力データに対応した異常検知モデルの構築を進め、完成予定の VS 取得デバイスによる実データを得て、性能評価を試みる。

E. 結論

バイタルサインの統合的評価のためのソフトウェア開発においては、昨年度選定した Acute Toxicity Vital Signs Score (ATVSS、仮称) の定義に利用可能なアルゴリズムのうち、教師あり深層学習について、判別性能の評価を行った結果、本研究の目標実現に有効な手法であることが判明した。

F. 研究発表

1. 論文発表

Nock R, Polouliakh N, Nielsen F, Oka K, R Connell C, Heimhofer C, Shibani K, Ghosh S, Aisaki K, Kitajima S, Kanno J, Akama T, Kitano

H: A Geometric Clustering Tool (AGCT) to robustly unravel the inner cluster structures of time-series gene expressions. PLoS One 2020; 15(7): e0233755.

[doi 10.1371/journal.pone.0233755]

2. 学会発表

高橋 祐次、種村 健太郎、相崎 健一、北嶋 聡、急性毒性試験の近代化によるテトロドキシンの中枢影響評価、第 47 回日本毒性学会学術年会 (2020.6.29.) オンライン

種村 健太郎、齊藤 洋克、古川 佑介、相崎 健一、北嶋 聡、菅野 純、低用量/低濃度化学物質の発生-発達期ばく露による情動認知行動毒性～情動認知行動毒性評価系の国際標準化に向けた対応～、第 47 回日本毒性学会学術年会 (2020.6.30.) オンライン

菅野 純、北嶋 聡、相崎 健一、小野 竜一、Percellome Project における精度管理とその解析への影響、第 47 回日本毒性学会学術年会 (2020.6.30.) オンライン

小野 竜一、相崎 健一、北嶋 聡、菅野 純、化学物質の反復投与によるゲノムワイドなヒストン修飾の変化、第 47 回日本毒性学会学術年会 (2020.6.30.) オンライン

相崎 健一、長谷 武志、北嶋 聡、小野 竜一、北野 宏明、菅野 純、Current and future application of PERCELLOME database as a part of big data to toxicological research、第 47 回日本毒性学会学術年会(2020.7.1.) オンライン

高橋 祐次、森田 紘一、辻 昌貴、菅 康佑、相崎 健一、大久保 佑亮、種村 健太郎、北嶋 聡、急性毒性試験の近代化による毒性機序研究、第 3 回医薬品毒性機序研究会 (2021.1.15) オンライン

G. 知的所有権の取得状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし