

令和4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による急性毒性試験の判定基準策定と代替法に資する研究
-診断学とAIによる致死性予測と人道的エンドポイントの設定-

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

バイタルサインの統合的評価による急性毒性試験の判定基準策定を主目的とした本研究班において、計測したバイタルサインの諸項目から「診断学」を基盤にした「概略の致死量」を推定するための統合的解析手法の開発と、これを実装したソフトウェアの開発を最終目標として研究を行った。今年度は、(1)バイタルサイン測定機器を限定することなく汎用性を持たせること、(2)学習プロセスを必要としないこと、の前提条件に適合し、時系列データ、特に心電図のような繰り返しパターンのある波形データの解析に有用な Matrix Profile アルゴリズムの導入を検討した。

A. 研究目的

バイタルサインの統合的評価による急性毒性試験の判定基準策定を主目的とした本研究班において、計測したバイタルサインの諸項目から「診断学」を基盤にした「概略の致死量」を推定するための統合的解析手法の開発と、これを実装したソフトウェアの開発を最終目標とする。

B. 研究方法

B-1 学習・評価用データ

バイタルサインデータとして想定される測定項目としては、血圧やパルスオキシメーターによる SpO₂、心電図などの、時系列データが想定されるため、評価用のバイタルサインデータとしては、主として心電

図データを用いた。ただし最終的にはバイタルサインの種類や測定機器を限定することなく汎用性を持たせることを重視して、バイタルサイン以外の時系列データも追加評価用として取り入れた。

B-2 解析計算及びソフトウェア生成

異常検知に利用し得る人工知能等の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.3.9.1)を使用した。汎用データ処理ライブラリとして Numpy (ver.1.19.5)、Pandas (ver.1.2.1)、データ可視化ライブラリとして Matplotlib (ver.3.3.4)を使用した。また Matrix Profile アルゴリズムの実装ライブラリとして matrixprofile (ver.1.1.10)を導入した。

Python スクリプト実行環境としては Visual Studio Code (ver.1.74.2), Jupyter Extension for VSCode (ver. v2022.11.1003412109) 或いは

Google Colaboratory を使用した。

B-3 計算精度検証

計算精度は必要に応じて 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かめた。

C. 研究結果

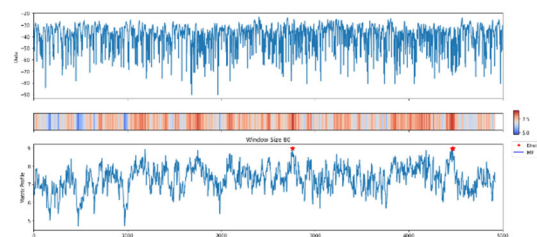
今年度は、時系列データ、特に心電図のような繰り返しパターンのある波形データの解析に有用な Matrix Profile (MP) アルゴリズムの性能評価や要件検討を実施した。

MP は基本的に、部分データ(A)と、時系列データ全体をAと同じ長さに分解した部分時系列データBの類似性を計算するものであり、異常検知だけでなく、繰り返しパターンの検出についても有用で、数理的なアルゴリズムであるため、事前学習や特別な処理を必要としない。またアルゴリズムの実装において優れたライブラリが提供されており、最小限のパラメータのみで実行可能である。波形データのベースラインの調整など、事前の正規化処理も不要であり、汎用性が高い。

性能評価としては、具体的には、ラベル付加や正規化などの前処理を適用していない、30 分間(測定頻度 650,000 回)連続測定した心電図データにおいて MP 解析を実行した。

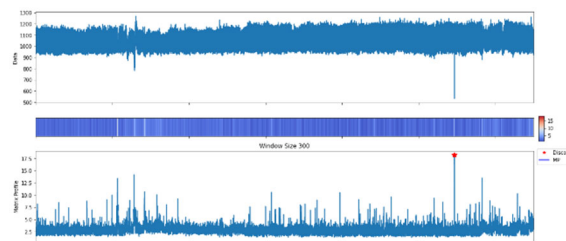
まず、MP にパラメータを一切与えず解析 (`matrixprofile.analyze`) を実行した場合、基本的な評価単位となる部分データのサイズ(幅)を一定の範囲で変化させて自動解析する(“Pan-Matrix Profile” と呼称)が自動実行されるが、データ点数が数千程度の小規模データを解析した場合を除き、実用的な時間内に計算が終わらず、また解析結果も十分な異常検知性能を示さなかった。

次に、時系列データ(波形)の目視確認を元に、部分データのサイズ(幅=周期的に出現するパターンのサイズ)を適切に指定したところ、その他のパラメータを一切指定せずとも自動的に解析が実行され、実用的な時間内に十分な異常検知性能を示した。



MatrixProfile ライブラリが自動生成する解析図。上段が心電図を示し、中段がヒートカラー、下段がピークにより、異常検知状況を示す。入力データは正規化していないためベースライン変動がそのまま残っている(上段)が、異常検知への影響は少ないこと(中段・下段)が示されている

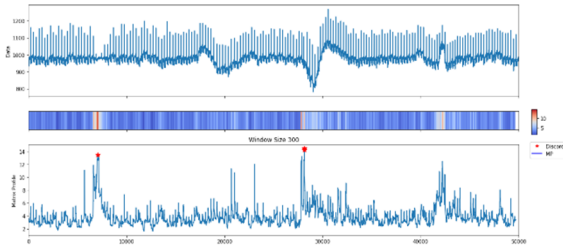
ただし、部分データに比べ、長大な時系列データを解析した場合は相応の実行時間が掛かること、また時系列データ内に極端な変動があった場合、異常の自動検知に大きな影響を与えてしまうこと、が確認された。



長大な時系列データを入力した場合、異常検知状況は不調(中段は青く、下段はピークが少ない、つまり候補が少ない)で、専ら大きな変動(artifact の可能性が高い)を異常と検知している)

また、心電図データと同様に周期性のある時系列データとして、一定の動作を繰り返す機械音を数値化したものを MP 解析したところ、心電図での解析結果と同様の結果を得た。これは同アルゴリズムが様々な種類の時系列データに対応可能であることを示唆している。

一方、規則性に乏しいランダムな時系列データや、不整脈時の心電図など異常パルスが有意に頻出しているようなデータを MP 解析すると、偶然の波形を繰り返しパターンと誤認したり、正常心電パルスを異常と検出したが、MP により自動出力された解析図を確認すれば容易にそれと認識できた。



周期性の乏しい時系列データを入力した場合、異常検知状況は候補の過剰検出状態(中段は赤く、下段は乱雑な小ピークしかない)で、有意かどうか判別困難な部分を異常と検知している。

D. 考察

先行研究では、汎用性の高い異常検知アルゴリズムとして人工知能(AI)、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機械学習に着目し、代表的な、畳み込みニューラルネットワーク(CNN: Convolutional Neural Networks)モデルや自己符号化器(AE: AutoEncoder)について、急性毒性試験においてバイタルサイン評価に用いる際の実用性と問題点を検討した結果、専門家による安全性評価に資する情報を提供し得ることを示した。一方でAI、特にCNNモデルにおいては事前の学習工程が出力データの精度に大きく影響することが確認され、学習のために十分な量の異常データのサンプリングが難しいバイタルサインの性質上、初期運用に際して困難が伴うことが予想された。

これらに比して、今年度取り上げたMatrix Profile(MP)は機械学習の類ではなく、基本的には数理的な手法といえるため、学習済みの機械学習モデルのように、検出された特徴波形(パターン)の分類を行う事は困難だが、原理的に事前の学習工程を必要としないため、汎用性、即応性に優れており、解析速度も急性毒性試験実施中のリアルタイム処理が可能なほどに高速だった。

この特性は、本研究班の目指す、急性毒性試験中の動物の死亡の回避(致死性予測)の実現において大変有効である。

具体的な方策としては、全ての測定を終えてから全測定範囲を一括解析するのではなく、急性毒性試験進行中に、現時点から遡って、高速処理が可能な範囲(データ量)の時系列データを対象にしたMP解析を連続実行することで、異常の発生をほぼリアルタイムに検知可能になると思われる。

不整脈心電図などの異常が多発しているデータを入力した際の誤識別については、むしろ人道的見地からそのような状態に至る前に急性毒性試験を終了すべきであり、適切な運用を行えば実用に際しての技術的問題にはならないと予想される。

今回の研究結果は、MPと機械学習のいずれが優れているか、というのではなく、組み合わせによる相乗効果を示唆するものと考えられる。実用化イメージとしては、急性毒性試験実施中はMPによる自動且つ高感度の致死性予測をその場でリアルタイムに行い、試験実施後の詳細解析や評価においては機械学習による特徴分類によって、専門家による判断の補助を行うなど、適切な使い分け・組み合わせが有効と考えられる。

またMPは、上記の性能に加え、①実行に際して設定が必要なパラメータがほとんどなく、唯一必要と思われる部分データの幅についても、設定に際して専門的な知識を前提としないため、データサイエンスの専門家でなくとも運用が容易であること、さらには②測定された時系列データをそのまま入力しても異常検知を実施できるため、バイタルサイン統合解析ソフトウェアの機械学習モジュールを測定機器や試験施設に最適化するために必要な学習用データを抽出・生成する際にも活用可能であること、の2点からも導入するに値するものと評価された。

E. 結論

今年度検討したMatrix Profileアルゴリズムは、最適化の進んでいない段階での異常検知において有用であり、また先行研究で評価した機械学習モデルとの組み合わせによる相乗効果が予想されることから、本研究班の目標実現に有効な手法であることが確認された。

今後も、バイタルサインの統合的解析に有効と思われる手法や追加項目を検討すると共に、従来からの統計指標等の手法や機械学習技術を組み合わせ、本研究班で開発したバイタルサイン取得デバイス等による多項目の入力データに対応する評価システム的设计・構築を進める。

F. 研究発表

1. 論文発表

相崎 健一, 小野 竜一, 菅野 純, 北嶋 聡:
Percellome プロジェクトへトランスクリプトミクスと
エピジェネティクスによる毒性分子機序の探求～.
日本薬理学雑誌, 2022; 157: 200-206.

2. 学会発表

菅野 純, 相崎 健一, 小野 竜一, 北嶋 聡: 新型反
復曝露トランスクリプトミクスから見た発癌
エピジェネティクスの考察. 第49回日本毒性
学会学術年会 (2022.6.30)

菅野 純, 相崎 健一, 小野 竜一, 北嶋 聡:
Percellome project からみた毒性AIの展望. 第
49回日本毒性学会学術年会(2022.7.2)

J. Kanno, K.-I. Aisaki, R. Ono, S. Kitajima:
Histone Modification, DNA Methylation, and
mRNA Expression Analysis of Murine Liver
Repeatedly Exposure to a Chemical. The
XVIth International Congress of Toxicology
(2022.9.19)

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