

機械学習とクラスタリング手法による複数種類の 正常値・異常値の同時分類技術

研究のポイント

- 水位の観測データに含まれる異常値を判定するため、自己組織化マップによる機械学習とクラスタリング手法を用いて複数種類の異常値と正常値を同時に分類する技術を開発しました。

研究の背景

- 近年ICTの普及により、大量のデータの収集が可能になりつつあるものの、その活用にあたってはデータの品質に課題があります。例えば、観測データには、計測機器の不具合等で発生する複数種類の異常値、通常の特徴と異なり異常値と区別が難しい正常値を有する複雑なデータがあり、それぞれの種類を明確かつ同時に認識できる検知技術が求められています。

技術の特徴と適用事例

- 農業水利施設等で観測される水位データには、ある値が急激に変化する「スパイクノイズ」や連続する値が全体的にシフトする「スライドずれ」等の異常値のほか、平常時の水位変化「正常値A」と洪水時の水位変化「洪水イベント」のように複数の正常値があります。本技術では、異常値と正常値を合計4つのラベルにより区別し、同時分類するものになります(図1)。
- 同時分類には、複数の値を1つの塊にしたデータ(ベクトル化)を、データ間の違いを距離で表現する自己組織化マップ(SOM)とSOMの要素情報を多数決でクラスタリングする多数決法を用いて、2次元マップ上に分類結果を可視化します。
- 異常値を多く含むように人工的に生成した時系列データを用いて、従来型のクラスタリング手法(K-means法)と多数決法で評価しました。ラベルの分類状態を2次元マップ上で可視化すれば、多数決法では正常値Aを分類した青色領域と正常値Aを表す青色ドット色が一致しており、分類精度が高いことが分かります(図2)。

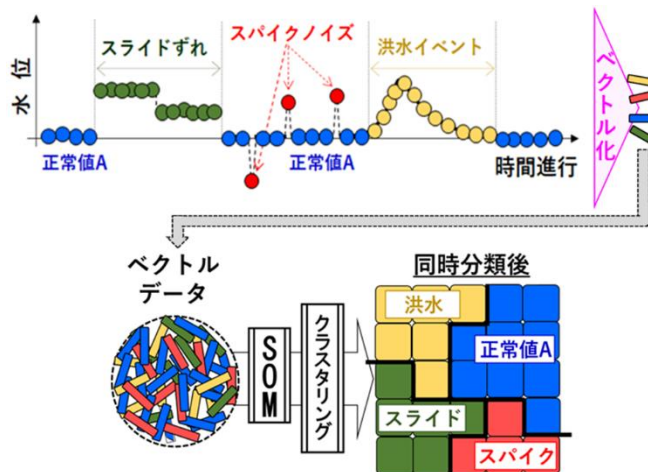


図1 複数種類の正常値・異常値の同時分類技術のイメージ

期待される活用例

- 本技術は、時系列データ全般に適用可能であり、複雑なパターンを有する測定データの品質を高めることが可能になります。例えば、心電図のような精密機器の稼働状況を計測する電気信号では、複数タイプの正常・異常パターンが含まれますが、それらを同時に検知することが可能になります。

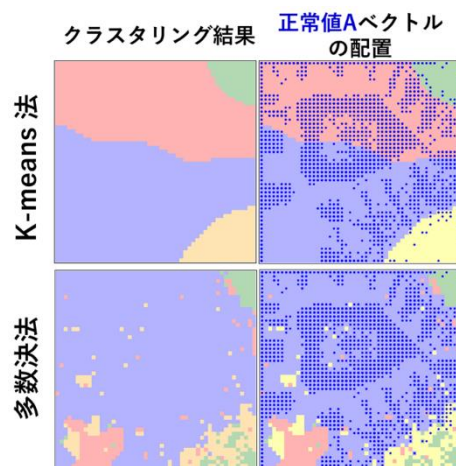


図2 K-means法と多数決法の分類結果(正常値Aのみの比較)
背景色=正常、洪水時などのデータの状態を4区分して緑、桃、紫、黄色で表示、ドット色=正常値を青色ドットで表示