

1. 氏名	小野 智司
2. 所属機関	鹿児島大学 大学院理工学研究科 情報生体システム工学専攻
3. 研究題目	異常検知や変化点検知における教師信号付き学習データの合成方式の提案
4. 研究の目的:	<p>本研究の目的は、系列データにおいて、専門家でも発見が困難な微小な変化や異常を検出できる技術を実現することである。一般に、異常や変化を検知する機械学習技術は、正常データと比較して異常や変化等のデータが圧倒的に少ないために、教師無し学習が用いられることが多い。このため、比較的大きな変動パターンでなければ変化や異常を検出することが難しい。</p> <p>これに対して本研究では、教師信号が付与されていないデータをもとに、教師信号付き学習データを合成し、教師有り学習を可能にする方法を提案する。教師有り学習を行うことで、従来は発見が困難であった自然変動に対してより微小な変動パターンのみを含む変化や異常であっても、検出が可能となる。</p> <p>本研究で開発する方式は、地球科学分野、医療分野、製造分野を始めとする幅広い分野に応用が可能である。地球科学分野の種々の計測における、長期的な気象変動の予測に使用される基礎データの整備に貢献する。その他、生体計測データによる患者の取り違え等の医療事故の発生の検知や、機械の早期異常発見等にも応用が期待できる。</p>

5. 研究の内容(手法、経過、評価など。書ききれない場合には、同一様式のページを追加してください。):

提案する教師信号付き学習データの合成は、複数の正常なデータを組み合わせることで行う。すなわち、2つの異なる系列データを、ある時点で分割し、組み替えることで新しい系列データを生成する。このとき、分割した時点で何らかの異常や変化が生じたとみなし、異常や変化があったという教師信号を付与できる(図1)。

例えば、気象観測データでは、変化を含むサンプルを、(拠点数)<sup>2</sup> × (時点数)<sup>2</sup> のオーダで合成することが可能となる。以上のようにして生成したデータをもとに学習器の訓練を行うことで注目すべき特徴を学習することが可能となり、微小な変動を検知することが可能となる。

生成した学習データは、任意の教師有り学習器で利用が可能であり、多様な分野のタスクに応用が可能である。本研究では、既に検証済の気象観測に加えて、海洋観測、医療、自然言語処理の各分野に提案する学習データ合成方式を応用し、その有効性を検証する。

図1 提案する教師信号付き訓練データの合成方式の外洋。気象観測に応用する場合は、(拠点数)<sup>2</sup> × (時点数)<sup>2</sup> オーダのサンプルを合成可能である。

6. 研究の成果と結論、今後の課題:

本研究では、提案する学習データ合成方式を、海洋観測、予防医療、ならびに自然言語処理の各分野に応用し、それぞれにおいて、海洋観測データの観測エラー発見、心拍データのみからの体長憎悪の判定、ならびに日本語文における妥当でない接続関係の検出を試みた。

海洋観測への応用では、リアルタイム海洋観測システム Argo による観測データに提案方式を適用し、Argo フロートによる観測エラーの検出成功率の改善を試みた。図 2 に結果を示す。グラフの合成方式 2 が提案方式によりデータ合成を行った後にニューラルネットワークにより観測エラーの結果を試みた際の F 値を示している。データ合成を行わない場合、および、合成方式 1 (従来のデータ合成手法 mixup)と比較して、特に浅層における観測エラーの発見性能が大幅に改善されることがわかる。

医療データへの応用では、集中治療室に入院中の患者データセットを用いて、心拍変動のみに着目して体調の悪化の検出が可能かどうかを試みた。特に、提案方式により、体調悪化を示す教師信号を用いない場合であっても体調悪化の検出が可能であるかを検討した。事例列データ用樹木モデルを識別に用いた場合の実験結果を表 1 に結果を示す。教師有り学習と比較して精度は低いものの、体調悪化の 30%以上の事例を検出できることから、大量のデータを活用できる状況であれば提案方式が有効である。

自然言語処理への応用では、論文等の文章において、文内のフレーズの接続関係が妥当でない箇所を検出するタスクに、提案手法を応用した。CiNii に掲載されている論文の概要をもとに作成したデータセットにおいて、被験者(学部 1 年次学生)よりも高い正解率で、接続関係が妥当でないフレーズを含む文章を発見できることを示した(表 2, 図 3)。

以上のような検証を通じて、提案方式は極めて単純な方式であるために多様な分野への応用が可能であることを示した。

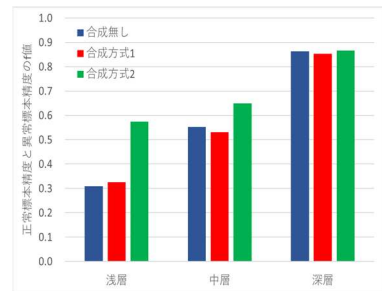


図 2 海洋観測データにおける結果. 浅層データにおいて観測エラーを検出性能が大幅に向上したことがわかる。

表 1 医療データにおける結果. 教師信号がないデータからであっても、体調が悪化したケースのうち 30%程度を検出できたことがわかる。

	正解率	再現率	適合率
教師有り	0.778	0.716	0.814
提案手法 (教師無し)	0.606	0.324	0.746

表 2 日本語文における接続妥当性の検証に応用した結果. 被験者よりも高い性能で、接続関係が妥当でない日本語文を検出できたことがわかる。

	正解率	F 値
被験者平均	0.638	0.541
提案手法(教師無し)	0.817	0.888

高圧縮率の画像符号化で生じやすい画質劣化を低減するために、▲Hough 変換は計算コストが最小二乗法と比べて大きいが発見率は堅牢である。

図 3 提案手法により検出された、接続

## 7. 成果の価値

### 7.1\_学術的価値:

本研究で提案したデータ構成方式は、近年注目されている自己教師あり学習の1つであるとみなすことができ、変化点検知を目的としている点で独創性が高い。また、従来は発見が困難であった微小な変化を検知できることから、変化点検知技術の適用範囲の拡大の貢献に寄与したものとする。

### 7.2\_社会的価値:

提案方式による学習データ合成は、教師信号が付与されていないデータに適用が可能であること、ならびに、微小な変化を検出できることから、機械等の故障診断、医療、環境計測における観測データ品質管理など、多様な領域での活用が見込まれる。

### 7.3\_研究成果:

#### 「研究論文(原著)」

[1] 小野智司: “進化計算を用いたパターン生成”, 画像電子学会誌, Vol.49, No.2, pp.154-159 (2020).

#### 「国際会議発表」

[2] O Daiki Tamashiro, Koji Yoshimitsu, Satoshi Ono: “An Attempt to Discriminate Health Condition Deterioration Using Heart Rate Variability by Machine Learning”, 28th International Symposium on Artificial Life and Robotics (2023).

[3] Kento Wakamatsu, Satoshi Ono: “TransUNet with Unified Focal Loss for Class-imbalanced Semantic Segmentation”, 28th International Symposium on Artificial Life and Robotics, 8th International Symposium on BioComplexity, 6th International Symposium on Swarm Behavior and Bio-Inspired Robotics (2023).

[4] Serina Higashi, Takahiro Kinoshita, Shigeki Hosoda, Ken-ichi Fukui, Satoshi Ono: “A Preliminary Study on Algo Float Trajectory Prediction using Multimodal Convolutional LSTM”, JpGU-AGU Joint Meeting 2020, No.ACG50-P04 (2020).

#### 「特許」

[5] O 小野智司, 前原 宗太郎, 福井 健一, 冨田 智彦: “教師情報付学習データ生成方法、機械学習方法、教師情報付学習データ生成システム及びプログラム”, 特許第 6989841 号 (2021.12.7 登録).

#### 「受賞」

[6] 学生会講演奨励賞, 電子情報通信学会九州支部 2022 年度(第 30 回)学生会講演会 (玉城 大生, 吉満 孝二, 小野智司: “機械学習による心拍変動指標を用いた体調増悪判別の試み”)