

令和 5 年 5 月 31 日現在

機関番号：17102

研究種目：基盤研究(B) (一般)

研究期間：2019～2022

課題番号：19H04176

研究課題名(和文) 製造インフォマティクスに向けた機械学習技術の開発と鉄鋼製造における評価

研究課題名(英文) Development of algorithms for manufacture informatics and its evaluation in steel industry

研究代表者

西郷 浩人 (saigo, hiroto)

九州大学・システム情報科学研究所・准教授

研究者番号：90586124

交付決定額(研究期間全体)：(直接経費) 12,400,000円

研究成果の概要(和文)：「高炉の異常検知問題」において、教師なし学習によるアプローチ(板倉(IBIS2022))と教師ある学習によるアプローチ(木崎 (IBIS2021))を開発した。CNNを用いた教師あり学習においては5分から15分前のデータの利用が精度の向上につながることを確認している。「マルチタスク学習による高温状態の粘度予測」のための方法を開発した(Saigo et al., Scientific Reports, 2022)。ロバストな外挿予測が出来るだけでなく、室温実験データを高温実験に役立てる転移学習法も提案した。

研究成果の学術的意義や社会的意義

教師なし学習は人手により教師ラベル作成の労力を減らすことを可能とする。現実世界の多くのデータは教師ラベルがないか、もしくはそのラベル付けに多大なコストが必要な場合が多いため、現実社会での実装において重要なテーマである。

一方で、機械学習手法の多くは過去のデータから学習し、その評価を交差検証に頼っているため、ロバストな外挿予測問題への取り組みは学術的に重要である。本研究では流体力学という現実の問題への解決策を示したものであり、同種の問題に適用可能である。

研究成果の概要(英文)：In the "Anomaly Detection in Blast Furnaces" problem, we have developed approaches using unsupervised learning based on the work of Itakura et al. (IBIS2022), and supervised learning based on the work of Kizaki (IBIS2021). In the supervised learning approach using CNN, we have confirmed that utilizing data from 5 to 15 minutes prior leads to improved accuracy.

We have also developed a method for "Viscosity Prediction of High-Temperature States through Multi-Task Learning" as described in the study by Saigo et al. (Scientific Reports, 2022). In addition to robust extrapolation prediction, we have proposed a transfer learning method that leverages room temperature experimental data for high-temperature experiments.

研究分野：機械学習

キーワード：機械学習 異常検知 鉄鋼生産 外挿予測 転移学習 ガウス過程 多相融体の粘度 深層学習

### 1. 研究開始当初の背景

鉄鋼生産は明治期以降の日本の産業活動の中心である。現在でも鉄鋼産業は日本経済を牽引する製造業の中で GDP3 位を占めており、鉄鋼は自動車に次いで輸出額 2 位という基幹輸出品目である。また自動車や電器製品以外にも、造船や橋梁といった構造物からガス管や建材などのインフラ基盤まで、現代の生活は鉄鋼製品に囲まれていることからその重要性を知ることができる。一方で、その製造に用いる設備に注目すると、1 基 1000 億円を超える溶鉱炉をはじめとして巨額な投資が必要であり、生産や操業の効率化を進めることによって得られるメリットは非常に大きい。しかしながら、操業の効率化や安定化を左右するいくつかの重要なプロセスにおける操業データは継続的に計測されて保存されているにもかかわらず、これらを利用した運用はオペレーターとよばれる熟練工の勘に頼っている状態で、必ずしも有効に活用されてはいなかった。

鉄鋼生産はその長い歴史にも関わらず、高温の高炉の内部状態の直接観測が難しいために、その仕組みの完全な理解がなされていなかったことにも原因である。そこで、本研究ではその仕組みの理解と制御に向けた機械学習手法の開発を目指したものである。

### 2. 研究の目的

高炉の内部状態を予測して制御するために、【1】操業時の異常検知問題、【2】高温多相融体の粘度予測 という 2 つの目標を定めた。

「【1】操業時の異常検知問題」について説明する。溶融炉の操業データは等間隔に設置された圧力センサーによって記録され行列あるいはテンソル形式で保存されている。ここで異常の前後のデータを比較すると、事故の直前にはかすかな兆候が現れることがわかっている。ただし、様々な状況や計測データを見て人間が判断するので、異常状態の検出精度にはばらつきがある。一方で、画像データの解析は近年の深層学習、特に畳み込みニューラルネットワーク(CNN) によって大きく進展している分野である。そこで、畳み込みニューラルネットによる早期異常検知システムの開発を行う。ただし、高炉の状態を人でラベル付けするにはコストがかかる。そこで、教師なし学習によるアプローチも検討した。

「【2】高温多相融体の粘度予測」について説明する。即ち、鉄鋼生産ではその精錬の際にスラグという物質が生じるが、生産の効率化のためにはスラグの粘度を小さくして効率良く鉄鋼から分離することが必要である。しかし、そもそも高温の溶融物(スラグ)の粘度の観測は物理的に困難である。しかも、溶融鉄は高温域において構造変化あるいは物理化学的变化を起こすことにより粘度が非線形に変化すると考えられている。そこで、提案者らは転移学習によるアプローチを考える。転移学習では、類似したタスクを同時に学習することにより、データの少ないタスクの予測精度を高めることが可能である。ここで、観測データについて考察すると、物理的制約等により、パラメータの範囲外のデータを取得できないことがあるが、予測モデルは学習されたドメインの外でも安定した挙動を示すことが求められる。機械学習モデルの評価に通常利用される交差検証は内挿問題であるため、外挿を考慮した予測モデルの構築が必要である。

### 3. 研究の方法

#### 【1】操業時の異常検知問題

##### 1) 教師あり学習によるアプローチ

k 近傍法、2DCNN、3DCNN を利用して、2 クラス分類問題における比較を行った。分類モデルの評価には Precision と Recall から計算される PRAUC を用いた。これは、負例(ここでは正常ラベルのこと)の数が多いためにおいても正確に精度を把握するためである。全てのモデルは時系列データであり、異常の時点から n 分間のデータを利用している。検証データを用いて 3DCNN の精度を調べたところ、そのピークは 5 から 10 にあることがわかった。このことから、時系列情報がモデルの精度を高めることが結論づけられた。全てのモデルの比較結果を表に示す。

	kNN	2DCNN (平均)	2DCNN (最後)	3DCNN
1 月データ	0.269 (k=29)	0.324	0.556	<b>0.917 (n=5)</b>
2 月データ	0.472 (k=17)	0.490	0.899	<b>0.957 (n=10)</b>

3DCNN がいずれのデータにおいても最高の性能を示していることがわかる。2DCNN よりも精度が高いことは、時系列情報を効率的に取り組んでいることを示唆している。また、kNN よりも精度が高いことは、特徴である圧力マップの全体（グローバル）よりも局所的（ローカル）な部分に異常検知のための情報が偏在していたことを示唆している。全体を通して 3DCNN による分類は False Negative が少なく Recall が高い傾向があった。

## 2) 教師なし学習によるアプローチ

代表的なクラスタリング法である k 平均法、PCA、KPCA を利用して異常検知性能の比較実験を行った。比較結果を表に示す。

	kmeans	PCA	KPCA
異常データあり	0.0488	0.151	<b>0.528</b>
異常データなし	0.486	0.174	<b>0.853</b>

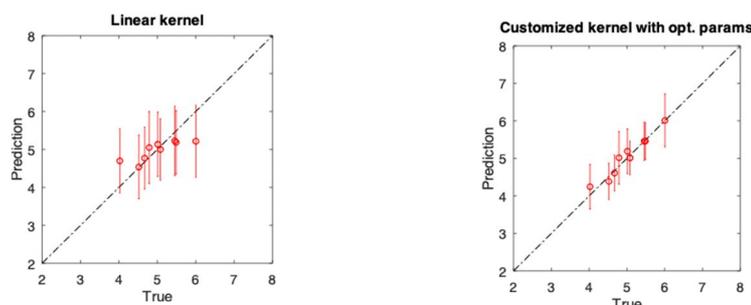
訓練データに異常データを含む場合と含まない場合を比べたところ、いずれの手法も前者の方が性能が高かった。また、手法としての精度はいずれの場合も KPCA が最高の性能を示した。PCA よりも KPCA の精度が高かったことは、異常検知に必要な特徴が非線形性を持つことを示唆している。

### 【2】高温多相融体の粘度予測

ロバストな外挿予測モデルを作成するために、流体力学の分野で利用される Einstein-Roscoe の式を事前知識として利用するガウス過程回帰を提案した。これを線形回帰(OLS)、LASSO、ランダムフォレスト(RF)、サポートベクターマシン(SVR)、マルチレイヤーパーセプトロン(MLP)と比較した結果を以下の表に示す。

	提案手法	OLS	LASSO	RF	SVR	MLP
内挿	0.075	0.00	0.037	0.13	0.189	0.32
外挿	<b>3.2</b>	301	10.0	12.0	14.0	15.0

内挿と外挿の設定で精度を比較したところ、提案手法の外挿における精度が際立って高く、事前知識を正確に取り入れられたことを確認できた。



上の図は、室温実験データを利用してパラメータを学習し、高温実験における精度の向上を目指した転移学習の結果である（左図は転移学習なし。右図は転移学習あり）。転移学習後は予測結果が対角線上に並び、予測精度が向上していることが分かる。また、予測の分散（丸の上下のバー）に注目すると、転移学習後は分散が小さくなり、予測の信頼度が高くなっていることがわかる。

## 4. 研究成果

### 【1】 操業時の異常検知問題

教師なし学習によるアプローチと教師あり学習によるアプローチを IBIS2022, IBIS2021 でそれぞれ発表した。他に例のない実データを用いた解析結果であるため独創性は高い。それぞれのアプローチを今後も研究する予定である。

### 【2】 高温多相融体の粘度予測

Einstein-Roscoe 式を事前知識として利用したロバストな外挿予測法を (Scientific Reports, 2022) にて発表した。比較的粘性の高い流体の解析におけるパイオニア的な研究であると考えており、産業の現場におけるインパクトも大きいと考えている。

## 5. 主な発表論文等

〔雑誌論文〕 計8件（うち査読付論文 8件/うち国際共著 8件/うちオープンアクセス 5件）

1. 著者名 Takayanagi, M., Tabei, Y., Suzuki, E., Saigo, H	4. 巻 9
2. 論文標題 Sparse nonnegative interaction models	5. 発行年 2021年
3. 雑誌名 IEEE Access	6. 最初と最後の頁 109994-110005
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2021.3099473	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 Li, W. and Saigo, H. and Tong, E. and Suzuki, E.	4. 巻 56(3)
2. 論文標題 Topic modeling for sequential documents based on hybrid inter-document topic dependency	5. 発行年 2021年
3. 雑誌名 Journal of Intelligent Information Systems	6. 最初と最後の頁 453-458
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Thapa, N., Chaudhari, M., McManus, S., Roy, K., Newman, R.H., Saigo, H., KC, D.B.	4. 巻 21(Suppl 3)
2. 論文標題 DeepSuccinylSite: a deep learning based approach for protein succinylation site prediction	5. 発行年 2020年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する
1. 著者名 Chaudhari, M., Thapa, N., S., Roy, K., Newman, R.H., Saigo, H., KC, D.B.	4. 巻 16
2. 論文標題 DeepRMethylSite: a deep learning based approach for prediction of arginine methylation sites in proteins	5. 発行年 2020年
3. 雑誌名 Molecular Omics	6. 最初と最後の頁 -
掲載論文のDOI（デジタルオブジェクト識別子） なし	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 該当する

1. 著者名 Al-barakati, H.J., Saigo, H., Newman, R.H., Bahadur, K.C.D.	4. 巻 15
2. 論文標題 RF-GlutarySite: a random forest predictor for glutarylation sites	5. 発行年 2019年
3. 雑誌名 Molecular Omics,	6. 最初と最後の頁 189-204
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12859-020-3342-z	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Al-barakati, H.J., Thapa, N., Saigo, H., Roy, K., Newman, R.H., Bahadur, K.C.D.	4. 巻 18
2. 論文標題 RF-MaloSite and DL-Malosite: Methods based on random forest and deep learning to identify malonylation sites	5. 発行年 2020年
3. 雑誌名 Computational and Structural Biotechnology Journal	6. 最初と最後の頁 852-860
掲載論文のDOI (デジタルオブジェクト識別子) 10.1016/j.csbj.2020.02.012	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する

1. 著者名 Saigo, H., Bahadur, K.C.D, Saito, N.	4. 巻 12
2. 論文標題 Einstein-Roscoe regression for the slag viscosity prediction problem in steelmaking	5. 発行年 2022年
3. 雑誌名 Scientific Reports,	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1038/s41598-022-10278-w	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

1. 著者名 Pratyush, P., Pokharel, S., Saigo, H., KC.D.B.	4. 巻 24(1)
2. 論文標題 pLMSNOSite: an ensemble-based approach for predicting protein S-nitrosylation sites by integrating supervised word embedding and embedding from pre-trained protein language model	5. 発行年 2023年
3. 雑誌名 BMC Bioinformatics	6. 最初と最後の頁 -
掲載論文のDOI (デジタルオブジェクト識別子) 10.1186/s12859-023-05164-9	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 該当する

〔学会発表〕 計12件（うち招待講演 0件 / うち国際学会 7件）

1. 発表者名 Yafune, R., Sakuma, D., Tabei, Y., Saito, N., Saigo, H.
2. 発表標題 Automatically mining relevant variable interactions via sparse Bayesian learning; International Conference of Pattern Recognition
3. 学会等名 International Conference for Pattern Recognition (ICPR) (国際学会)
4. 発表年 2021年

1. 発表者名 木崎亮介, 西郷浩人
2. 発表標題 深層学習を利用した高炉内の異常検知
3. 学会等名 第24回情報論的学習理論ワークショップ(IBIS)
4. 発表年 2021年

1. 発表者名 木崎亮介, 西郷浩人
2. 発表標題 深層学習を利用した高炉内の異常検知
3. 学会等名 人工知能学会 第116回人工知能基本問題研究会(SIG-FPAI)
4. 発表年 2021年

1. 発表者名 Li, W., Matsukawa, T., Saigo, H., Suzuki
2. 発表標題 Context-Aware Latent Dirichlet Allocation for Topic Segmentation
3. 学会等名 Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD 2020) (国際学会)
4. 発表年 2020年

1. 発表者名 Yafune, R., Sakuma, D., Tabei, Y., Saito, N., Saigo, H.
2. 発表標題 Automatically mining relevant variable interactions via sparse Bayesian learning; International Conference of Pattern Recognition
3. 学会等名 International Conference of Pattern Recognition (ICPR2020) (国際学会)
4. 発表年 2021年

1. 発表者名 Thapa, N., Chaudhari, M., McManus, S., Roy, K., Newman, R.H., Saigo, H., KC, D.B.
2. 発表標題 DeepSuccinylSite: a deep learning based approach for protein succinylation site prediction
3. 学会等名 Joint GIW/ABACBS-2019 Bioinformatics Conference (国際学会)
4. 発表年 2019年

1. 発表者名 Hussam Albarakati, Hiroto Saigo, Robert Newman and Dukka KC
2. 発表標題 SVM-GlutarySite: A Support Vector Machine-Based Prediction of Glutarylation Sites from Protein Sequences
3. 学会等名 MCBIOS2019 (国際学会)
4. 発表年 2019年

1. 発表者名 矢船 僚一朗、西郷 浩人
2. 発表標題 アイテムセットを用いたスパースベイズ学習
3. 学会等名 人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 高柳 未来、田部井 靖生、西郷 浩人
2. 発表標題 変数間作用を考慮した非負スパースモデルの正則化経路探索
3. 学会等名 人工知能学会全国大会
4. 発表年 2019年

1. 発表者名 Ryoichiro Yafune, Daisuke Sakuma, Yasuo Tabei, Noritaka Saito, Einoshin Suzuki and Hiroto Saigo
2. 発表標題 A Sparse Bayesian Approach to Combinatorial Feature Selection and Its Applications to Biological Data
3. 学会等名 ICBBB2020 (国際学会)
4. 発表年 2020年

1. 発表者名 Kohei Oyamada and Hiroto Saigo
2. 発表標題 Bayesian Optimization for Sequence Data
3. 学会等名 ICBBB2020 (国際学会)
4. 発表年 2020年

1. 発表者名 板倉健大; 西郷浩人
2. 発表標題 機械学習・深層学習を用いた高炉の教師なし異常検知
3. 学会等名 第25回情報論的学習理論ワークショップ(IBIS)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

-

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
研究 分担 者	齊藤 敬高  (Saito Noritaka)  (80432855)	九州大学・工学研究院・准教授    (17102)	

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------