様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

令和　５ 年　 ６ 月　 ２ 日現在

機関番号：１２６０１
研究種目：基盤研究(C)（一般）
研究期間：2019～2022
課題番号：１９Ｋ１２０９４

研究課題名（和文）データ駆動健全性監視のための転移学習と説明能力に関する研究

研究課題名（英文）Study on transfer learning and explainability of data-driven health monitoring

研究代表者
矢入　健久（Yairi, Takehisa）

東京大学・先端科学技術研究センター・教授

研究者番号：９０３１３１８９

交付決定額（研究期間全体）：（直接経費）　　3,300,000円

研究成果の概要（和文）：本研究は、データ駆動型の健全性監視法が抱える2つの問題の解決に取り組んだ。第一の問題は、現実の人工システムにおいては事前に十分な量と質を兼ねた訓練データを入手することが困難または非常に高価であること、第二の問題は、機械学習により帰納的に得られたモデルが対象人工システムのドメイン知識と乖離しているために実用に耐える説明性を担保していないことである。本研究ではこれら2つの問題に対して、工学者・専門家にとって解釈性の高い潜在変数-状態空間モデルと最新の機械学習手法との融合を図ることによって、ドメイン知識の活用による必要訓練データ量の削減と、データ駆動健全性監視の説明性の向上を実現した。

研究成果の学術的意義や社会的意義
人工知能・機械学習はインフラや生産システムなどが安全かつ効率的に運用されているかどうかを監視する目的においても大いに期待されているが、学習に膨大な訓練データが必要であること、および、モデルがブラックボックスになり説明性に欠けることが大きな懸念事項である。本研究は、最新の機械学習と伝統的な状態空間・潜在空間モデルを統合して動的なシステムのモデルを学習する方法を開発することによって、これらの問題の解決に貢献した。

研究成果の概要（英文）：This study tackled two issues associated with data-driven health monitoring methods for artificial systems. The first issue is the difficulty or high cost of obtaining a sufficient amount and quality of training data in real-world artificial systems. The second issue is the lack of explanatory power in data-driven models obtained through machine learning, as they diverge from the domain knowledge of the target artificial system, making them less practical. In this study, we addressed these two problems by integrating interpretable latent variable-state space models, which are highly interpretable for engineers and experts, with state-of-the-art machine learning techniques. This integration enabled us to reduce the required amount of training data through the utilization of domain knowledge and improve the explanatory power of data-driven health monitoring.

研究分野：健全性管理

キーワード：データ駆動型健全性管理　機械学習　動的システム学習　異常検知

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

本研究課題を開始した当初の背景として、以下の３点が挙げられる。

【背景 1】　大規模人工システムのためのデータ駆動健全性監視技術への期待と不安
現代社会は、電力、交通、生産などの様々な産業領域において多種多様の大規模人工システムが稼働することによって維持されている。従ってそれらのシステムの健全性を監視し保全することが極めて重要である。近年、計測通信技術の発展によって、あらゆる領域で高頻度・高次元の計測データが入手可能になり、データ駆動型の健全性監視技術に大きな期待が集まっている。実際、データ駆動型健全性監視の基本的フレームワークは既に確立されていると言って良い。しかし、データさえ豊富にあれば従来の知識駆動型の健全性監視で利用されていた専門家知識が不要になるかどうかは明らかでなく、むしろ懐疑的な意見が強い。

【背景 2】 教師なし学習に基づくデータ駆動健全性監視における「訓練データ不足」問題
データ駆動型健全性監視の基本的な考えは、正常データに対して教師なし学習手法を適用することによってシステムの正常モデルを帰納的に獲得することである。従って、起こり得る全ての正常状態におけるデータが必要となるが、現実的には、１つ１つの人工システムに対してそのような理想的な訓練データを用意することは極めて困難である。その結果、異常を誤検知することが深刻な問題になっている。

【背景 3】 データ駆動健全性監視における説明性の問題
近年、機械学習では深層学習技術の発展が著しく、教師なし学習の分野においても変分オートエンコーダや敵対的生成ネットワークなどの手法が開発され、データ駆動健全性監視への応用も報告されている。しかし、これらの方法では、異常が検知されたときに、それがシステムのどの部分で発生し、どれだけ深刻であるかについての説明性が欠如している点が指摘されている。これは、大規模人工システムの健全性監視においては重大な問題である。

２．研究の目的

本研究は上で述べた背景を踏まえ、大規模人工システムのためのデータ駆動型健全性監視において、(1) 転移学習法を明らかにすることによって訓練データ不足問題を解決することと、(2) 潜在変数－状態空間モデルを導入することにより説明性を実現することを目的とする。具体的には以下の通りである。

【目的 1】データ駆動健全性監視のための事前知識の利用法と転移学習法の確立
訓練データ不足を補うために、監視対象のシステム挙動に関する事前知識をモデル学習に取り込む方法を明らかにすることが本研究の第一の目的である。また、教師なし学習に基づくデータ駆動監視を対象とした転移学習法を開発することによって、個々のシステムでは訓練データが不足している場合でも、他の関連システムのデータを利用して正常モデルを学習する方法を明らかにする。

【目的 2】潜在変数-状態空間モデルに基づくデータ駆動健全性監視の説明性実現
観測対象である人工システムから得られる多次元時系列データに対して深層学習のような非線形性と自由度の高いモデルを素朴に適用することは、学習結果をブラックボックス化し説明性を損なう。本研究では、潜在変数群とそれらのダイナミクス構造、すなわち、状態空間モデルをある種の制約として課すことによって、可視性・説明性を実現する。

３．研究の方法

本研究は、データ駆動型の健全性監視の対象となる人工システム群に共通する特性として、内部状態の時間遷移に関して規則性や制約条件、微分方程式などの形で事前知識が得られることに着目し、そのダイナミクスを潜在変数・状態空間モデルによって表現することによって、データから推定しなければいけないパラメータ数の削減と、システムの挙動に関する説明性の向上を図る。具体的には、

【方法１】従来のデータ駆動型アプローチのようにシステムの正常挙動をすべてデータから学習するのではなく、システムの潜在・内部状態に関する遷移法則や微分方程式、制約条件等の専門知識を利用することによって比較的少量の訓練データに対しても汎化性能の確保を図る。また、システムモデルを内部状態のダイナミクスに関する部分と、計測・観測データの生成部分と

に分け、前者を同種のシステム間で転用することによって、効率的な転移学習の実現を図る。

【方法２】従来のデータ駆動型の健全性監視ではデータの再構成・予測誤差に基づく異常検知が主流であるのに対して、本研究ではシステムの内部・潜在状態を推定することによって、異常検知時や意思決定時の説明性の向上を図る。

４．研究成果

(1) 非線形偏微分方程式系のデータ駆動型推定

流体など高次元かつ非線形のシステムの挙動を純粋なデータ駆動型のアプローチによって同定するためには膨大な訓練データが必要になる。それに対して本研究では、対象となるシステムの潜在空間において部分既知の偏微分方程式が成立していると仮定し、その偏微分方程式を深層ニューラルネットワークによってデータから近似的に推定する方法を提案し、比較的少量の訓練データでも汎化能力の高いモデルを学習できることを示した。

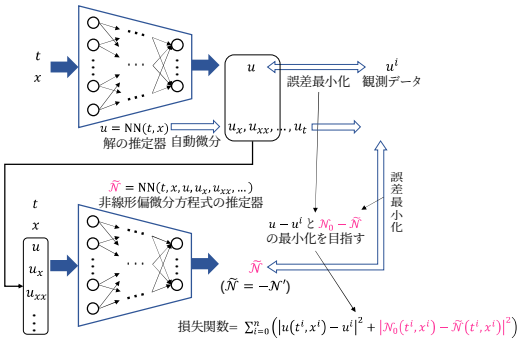

図 1　偏微分方程式系のデータ駆動型推定

(2) 状態空間モデルと深層ニューラルネットワークを用いた動的システム学習

状態空間モデルにおいて、状態遷移モデルには部分既知のダイナミクスモデルを利用し、観測モデルには、生成的な深層ニューラルネットワーク（変分自己符号器）を採用し、データからモデルパラメータを学習することによって、自然法則やシステム固有のダイナミクスに関する事前知識をモデルに埋め込みつつ、動画像のような超高次元時系列データ入力からシステムの内部状態遷移を推定・監視する手法を提案した。また、この手法を協調搬送ロボットの行動を模擬した画像時系列に適用し、複数種類の異常事象を的確に分類しつつ検知できることを実証した。

(3) 高次元時系列データの潜在変数空間への埋め込みによる解釈性の実現

変分自己符号器を用いた非線形次元削減と自己教師あり学習を組み合わせ、地球周回の人工衛星に本来備わる周期性を事前知識として利用することによって、潜在変数空間の各次元が解釈可能になり得ることを示した。

(4) 強化学習による意思決定の説明性に関する研究

室内移動ロボットのナビゲーションを題材として、強化学習によって移動方策を学習したエージェントが、学習後に各時刻ステップにおける意思決定の根拠を自然言語によって説明する方法を提案し、説明機能がユーザーである人間からの信頼性を向上させることを示した。

(5) 深層ニューラルネットワーク動的システムモデルを用いた状態・観測予測

上記研究成果(2)を発展させて、力学的運動に関する事前知識を埋め込んだ状態空間モデルと半教師あり変分自己符号器を融合することにより、形状モデルや画像特徴抽出を必要とせずに、宇宙デブリの３次元回転画像列から対象物体の姿勢状態を推定し将来の観測（見え方）を予測する手法を開発した。

５．主な発表論文等

〔雑誌論文〕　計4件（うち査読付論文　4件／うち国際共著　1件／うちオープンアクセス　1件）

１．著者名 KARINO Hidekazu、YAIRI Takehisa、NINOMIYA Tetsujiro、HORI Koichi	４．巻 63
２．論文標題 Estimating Aerodynamic Coefficients from Uncertain Data of D-SEND Aircraft with Gaussian Process Regression	５．発行年 2020年
３．雑誌名 TRANSACTIONS OF THE JAPAN SOCIETY FOR AERONAUTICAL AND SPACE SCIENCES	６．最初と最後の頁 257～264
掲載論文のDOI（デジタルオブジェクト識別子） 10.2322/tjsass.63.257	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 －

１．著者名 Anzai Yoshiyuki、Yairi Takehisa、Takeishi Naoya、Tsuda Yuichi、Ogawa Naoko	４．巻 4
２．論文標題 Visual localization for asteroid touchdown operation based on local image features	５．発行年 2020年
３．雑誌名 Astrodynamics	６．最初と最後の頁 149～161
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s42064-020-0075-8	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 －

１．著者名 Khan Samir、Liew Chun Fui、Yairi Takehisa、McWilliam Richard	４．巻 83
２．論文標題 Unsupervised anomaly detection in unmanned aerial vehicles	５．発行年 2019年
３．雑誌名 Applied Soft Computing	６．最初と最後の頁 105650～105650
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.asoc.2019.105650	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 Phong Nguyen X.、Tran Tho H.、Pham Nguyen B.、Do Dung N.、Yairi Takehisa	４．巻 10
２．論文標題 Human Language Explanation for a Decision Making Agent via Automated Rationale Generation	５．発行年 2022年
３．雑誌名 IEEE Access	６．最初と最後の頁 110727～110741
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/ACCESS.2022.3214323	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 －

〔学会発表〕　計5件（うち招待講演　1件／うち国際学会　3件）

１．発表者名
Ryosuke TANIDA, Chun Fui LIEW, Takehisa YAIRI and Yusuke FUKUSHIMA

２．発表標題
Dictionary Learning on Satellite Housekeeping Data: Profiling, Imputation, Novelty Detection

３．学会等名
33rd International Symposium on Space Technology and Science（国際学会）

４．発表年
2022年

１．発表者名
Khan, S., Yairi T

２．発表標題
Diagnosing intermittent faults through non-linear analysis

３．学会等名
21st IFAC World Congress 2020（国際学会）

４．発表年
2020年

１．発表者名
矢入健久

２．発表標題
動的システム学習の概要と研究動向

３．学会等名
第８回　制御部門マルチシンポジウム(MSCS2021) OS「データサイエンス×システム同定による制御技術の新たな発展」（招待講演）

４．発表年
2021年

１．発表者名
Riku Sasaki, Naoya Takeishi, Takehisa Yairi, and Koichi Hori

２．発表標題
Neural Gray-Box Identification of Nonlinear Partial Differential Equations

３．学会等名
16th Pacific Rim International Conference on Artificial Intelligence（国際学会）

４．発表年
2019年

１．発表者名 二木浩司，矢入健久
２．発表標題 変分オートエンコーダによる物体画像列からの姿勢推定
３．学会等名 第65回自動制御連合講演会
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
研究協力者	二木　浩司 (Hiroshi Futatsugi)		
研究協力者	グゥエン　フォン (Nguyen Phong)		
研究協力者	谷田　遼典 (Tanida Ryosuke)		
研究協力者	森　研人 (Mori Kento)		
研究協力者	安部　謙太朗 (Abe Kentaro)		

６．研究組織（つづき）

	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
研究協力者	笹木　陸 (Sasaki Riku)		

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