

令和 4 年 6 月 20 日現在

機関番号：12701

研究種目：基盤研究(B)（一般）

研究期間：2019～2021

課題番号：19H02159

研究課題名（和文）システム知の循環にもとづく受容可能な知能システムの創生

研究課題名（英文）Creation of the acceptable intelligent system based on the system intelligence circulation

研究代表者

濱上 知樹（HAMAGAMI, TOMOKI）

横浜国立大学・大学院工学研究院・教授

研究者番号：30334204

交付決定額（研究期間全体）：（直接経費） 13,200,000円

研究成果の概要（和文）：「受容可能な知能システムのためのシステム知の循環機構とこれを用いたシステム創生」にむけて主に4つの成果を得た。まず、専門家からの知識抽出の蒸留と共通および特有知識分離手法として、適応的閾値ブースティングを実現した。2つ目にマルチドメイン空間における機械学習バリュエーションモデルの高度化にむけて、マルチエージェントによるマルチドメイン空間の適応的分割と再利用の方法を明らかにした。3つめに、形状設計と改良のプロセスに機械学習による可視化フィードバックと説明性の向上をはかった。さらに階層型強化学習による制御問題の実現とその基本コンポーネントの抽出による知的構造の再利用と循環を実現した。

研究成果の学術的意義や社会的意義

サイバーフィジカルシステム(Cyber Physical Systems: CPS)は、今後の社会システムの姿を大きく変えるパラダイムである。CPSと人工知能の融合のためには、知能システム自らが、システム知を分析・説明・制御する仕組みを持ち、この仕組みの循環を通して、XAIを超えた受容可能な知能システム(Acceptable Intelligent Systems: AIS)へと発展することが必要である。本研究はそのために必要なシステム知の抽出と説明法、設計と実装法、応用方法について、要素技術から具体的アプリケーションまで実現したことで、本分野の学術的および社会的に大きな意義を有する。

研究成果の概要（英文）：These studies have achieved four main results toward the circulation mechanism of system knowledge for acceptable intelligent systems and system creation. First, we realized “adaptive threshold boosting” as a method for distilling knowledge extraction from experts and separating common and specific knowledge. Second, we clarified a method for adaptive partitioning and transferring multi-domain spaces by multi-agent approach for the advancement of machine learning valuation models in multi-domain spaces. Third, machine learning provides visual feedback and improved explanatory power to the shape design and refinement process. Furthermore, we realized the control problem by hierarchical reinforcement learning, and reused and circularized the intellectual structure by extracting its basic components.

研究分野：知能システム

キーワード：知能システム 機械学習 人工知能

様式 C - 19、F - 19 - 1、Z - 19 (共通)

1. 研究開始当初の背景

現実世界の様々な事象をリアルタイムに把握・集約し、ネットワークを介した大規模な計算・分析・予測処理にもとづいて現実世界の制御を最適化するしくみ サイバーフィジカルシステム(Cyber Physical Systems: CPS)は、今後の社会システムの姿を大きく変えるパラダイムとなった。交通、電力、水道をはじめとする社会インフラや、インダストリー4.0に代表される製造・生産、農業や水産などの1次産業、教育や医療などのサービスに至るまで、あらゆる問題領域(ドメイン)においてCPSによる革新的な効率化・最適化が期待できる。

高度なCPSの実現のためには、システム工学と機械学習を融合した知能システムエンジニアリングの高度化が必要である。知能システムの高度化に資する技術の中でも、大量のデータからEnd-to-Endの学習を行う深層学習は、複雑なシステムの構築を自動化する手段として期待されている。また、システム内部に埋め込まれた深層学習が獲得する構造には、複雑なシステムの理解と検証に必要な多くの知識が蓄積されていく。この知識を活用することで、システムの理解・検証のみならず、最適化や再利用への展開が可能になる。このように、深層学習は知能システムの分析と構築の両方にとって有望な要素技術である。

一方、深層学習が獲得する知識は、時として「暗黒知識(dark knowledge)」(Hinton2016)と呼ばれる極めて複雑な構造になることが知られている。暗黒知識の塊になった深層学習は、特定ドメインに対するEnd-to-Endの性能向上には役立つが、その結果を「検証」し「理解」するための了解性・解釈性を持たない。この問題は、深層学習のブラックボックス問題として知られているが、複雑システム的设计においてこの問題はさらに深刻となる。なぜなら、システムの公正性と正確性を担保するためには、複雑システム内の複数ドメインの振る舞いに関する知識 システム知(System Intelligence)が正確に分析・表現され、その妥当性が常に説明され制御可能であることが求められるからである。しかし、深層学習を用いて複雑化した知能システムは、そこに内在するシステム知を説明・利用する手段がないために、機能仕様は満たせても、これを制御し、保守・改良していくことは困難である。これが社会実装上の障壁となり、知能システムは、設計者・運用者にとって受容困難なものとなっている。

以上の背景のもと、説明可能なAI(eXplainable AI :XAI)の考え方を拡張し、CPSのような複数ドメインにまたがる複雑なシステムに対する蒸留(システム蒸留)を端緒とする、システム知の循環に基づく知能システムの実現という発想に至った。すなわち、知能システム自らが、システム知を分析・説明・制御する仕組みを持ち、この仕組みの循環を通して、XAIを超えた受容可能な知能システム(Acceptable Intelligent Systems: AIS)へと発展するという着想である。ここで「受容可能」とは、複雑なシステムのライフサイクル全体が、複数ドメインのシステム知によって検証可能になっていることを意味する。

2. 研究の目的

以上の背景のもと、本研究では、「受容可能な知能システムのためのシステム知の循環機構とこれを用いたシステム創生」という目標に向けて、以下の3つの小目標を設定した。

複雑システムの内部にある深層学習の学習結果(教師モデル)から、複数ドメインにわたる知識(生徒モデル)を蒸留する「システム知の分析法」を明らかにする。

- 1) 分析された複数ドメイン知識を学習し、ドメイン知識に基づく振る舞いを表現する「システム知の説明法」を明らかにする。
- 2) 分析・説明された複数ドメインのシステム知から、新たなシステム知を操作する「システム知の制御法」を明らかにする。
- 3) 受容可能な知能システムの応用例を医療・社会システムを例に実装する。

3. 研究の方法

2で述べた3つの小目標にむけてそれぞれ以下の方法で研究を進めた

- 1) システム知の分析法(アンサンブル蒸留)を明らかにする。実問題において知の再利用を阻む問題に、マルチドメイン間の知識利用がある。そのような場合でも有効に蒸留を行う方法として、学習対象のドメインに対して異なる正則化を施すことで、ドメインごとの教師モデルを抽出することで、複数の知の集合として蒸留をほどこすアンサンブル蒸留法を実現する。
- 2) システム知の説明法と制御法(Deep-Shallow ハイブリッド学習)を明らかにする。非線形関数が入れ子となる深層学習を説明可能にするために、Deepな学習機構で得られた教師モデルを、Shallowな学習機構を持つ生徒モデルで蒸留することにより「機能の学習結果」から「システム知の説明」への転移を実現する。逆にShallowな構造からDeepな構造へと転移させることで「システム知の説明」から「システム知の制御」への転移を実現する。
- 3) 以上の2つの小目標の応用をはかるために、具体的なアプリケーション・実問題に対し、システム知の循環機構の実現をはかる。具体的には車両自動制御、マルチドメインバリエーションモデル、FPGA ネットワーク自動設計、船舶設計問題、ケアプラン作成、生殖医療支援など、従来人の経験値に支えられていた高度な応用分野で、本研究の有能性を実証する。

4. 研究成果

(1) システム知の分析に関する検討 暗黙知の獲得アルゴリズム 参考文献

特定少数の人に関する知識抽出,特に専門家からの知識抽出に関しては,大きく2つの課題が存在している。人の持つ知識に依る意思決定の方法を情報空間から判断空間への写像と捉えると,専門家の知識は情報空間の分解能が高くなる場合が多い。その結果として,専門家の知識は非専門家と比較して高次に複雑になり,機械学習モデルによる知識の関数近似は過学習を起こしやすくなる。また,専門家の知識空間には偏りが存在する。その結果,複数人の専門家の知識を単一の関数で近似することが困難になる。特に,専門家の数が少なくそれらの知識に偏りが生じている場合には,専門家間で共通の知識を関数近似する際にバイアスの影響が無視できなくなる。本研究ではこれら2つの課題を,勾配ブースティングを用いた知識の関数近似により解決した。勾配ブースティングは複数の弱学習器から強学習器を推定する手法であり,次の理由から知識の関数近似に適している。まず,弱学習器の複雑さを調整することで情報空間の分解能を調整可能である。さらに,弱学習器の訓練時にそれらをグループ分けすることで,専門家の知識の個人差を分離することが可能である。こうした勾配ブースティングの特徴を活用し,「情報空間の絞り込みを行う手法」と「専門家の知識を共通知識と特知知識に分割する手法」を新たに提案する。この2つの提案手法に関して人工データと実データを用いて評価を行い,従来手法に比べ高い知識抽出性能が得られることを明らかにした。また,本手法を「高度生殖補助医療のための精子選別支援」システムに応用し,精子選別における専門家の暗黙知抽出評価を行った。その結果,特定少数の専門家に関する暗黙知獲得において本研究の有効性が確認された。

(2) システム知の分析に関する検討 マルチドメイン空間における機械学習バリエーションモデルの高度化 参考文献

バリエーションモデルは,様々な対象の時価,担保価値,与信等においてその価値を算定するための仕組みであり,従来は専門家による経験的な知識が不可欠な技術であった。機械学習の応用も図られてきたが,従来はデータが均一で同質であるシングルドメインの前提であった。しかし,複雑化し変化の早い世界にあっては,異質なデータが含まれたマルチドメイン空間を扱うことも一般的になりつつあり,従来の機械学習の前提では適切な推論ができなくなってきている。本研究では,バリエーションモデルをマルチドメイン空間に適応させるための方法について二つの手法を明らかにした。一つは適応的なドメイン分割方法である。対象の空間がマルチドメイン空間である場合バリエーションモデルが持つ市場性の原理に即し,マルチエージェントの仕組みを活用することで,ある対象データの周辺データからドメインを分割し,そのコンセンサスを推論に活用する手法を提案した。もう一つは,あるドメインから別のドメインへ知識を転移させる転移学習である。バリエーションモデルで扱うものの価値の評価には時間が経過して初めて明らかになる性質があるため,短期間で大量の教師データを生成することが非常に困難である。本研究では,ベイジアンネットを用いて対象ドメインから本質的な特徴量を抽出し学習に活用する手法を提案し,属人的で非柔軟な従来のバリエーションモデルが,汎用的で客観的なモデルへ高度化させるための技術的な進歩が得られた。

(3) システム知の説明法に関する検討 データ分布を考慮した深層学習による高精度推定法(船型データの分布を考慮した造波抵抗推定) 参考文献

近年の深層学習研究の進歩に伴い,様々な産業・分野で end-to-end 学習を用いた応用が広がっている。しかし,設計問題をはじめとする実データにおいては end-to-end の学習を直接行うことが困難な場面が多く存在する。特に,データのサンプリング密度が不均一である場合やインバランスが存在している場合,学習器が不安定になることが知られている。本研究では,これらの問題を解決するために,3次元形状の低次特徴量の抽出法と優先的選択的サンプリングによる偏りの解消をめざした。前者により,DAE の中間層の変化から周波数基底成分を取り出すことで3次元形状の特徴を効率的に表現する方法を明らかにした。特に,得られた結果を Grad-CAM によって可視化することで,設計に寄与する形状の特徴や,設計変更に伴う影響の範囲を明示することに成功し,推定の根拠と物理現象との対応について説明可能な手法を実現した。また後者により,偏りのあるデータ集合からロバストな学習結果を得る2段階の優先サンプリング手法を明らかにした。この方法によって,データの偏りと精度の関係を定量的に評価できるようになり,推定精度の信頼性を設計者に説明することが可能になった。さらに本手法を,船型形状設計問題に応用し,従来手法に比べて推定精度の向上が実現され,船型形状から造波抵抗を直接回帰する end-to-end のシステムの構築に成功した。

(4) システム知の説明法に関する検討 階層型強化学習と方策蒸留を用いた車両速度追従制御 参考文献

学習に必要なサンプル数の削減を目的とし,階層型強化学習と方策蒸留を用いた速度追従制御を提案した。自動車の試験走行のためのドライブロボットの従来手法として強化学習による制御があるが,方策の学習に多くの試行が必要である。提案手法では,State Abstraction を導入した階層型強化学習を用いて,異なる車両間の共通知識を下位方策として事前学習し,未知の車両に対してその知識を転移することで,個々の知的制御の説明原理を明らかにした。また,知識の転移だけでは,速度追従精度の改善に限界がある。そこで,方策蒸留を用いて方策構造を非階層型に変更し,非階層型強化学習により追加学習を行うことで,さらに精度を向上させる。実験を通して提案手法の有効性を確認し,サンプル数の削減が可能であることを明らかにした。

<引用文献>

Sasaki Hayato, Yamamoto Mizuki, Takeshima Teppei, Yumura Yasushi, Hamagami Tomoki, "Sperm Detection with Robustness to Overlap of Distributions by using Adaptive Thresholded Boosting", IEEJ Transactions on Electronics, Information and Systems Vol.139(12), pp.1461-1467, 2019.

Iwai Koichi, Akiyoshi Masanori, Hamagami Tomoki, "Bayesian Network Oriented Transfer Learning Method for Credit Scoring Model", IEEJ Transactions on Electrical and Electronic Engineering Vol.16(9), pp.1195-1202, 2021.

Li Xin, Arai Hiroshi, Hamagami Tomoki, "Wave-making Resistance Estimation Through Deep Learning Considering the Distribution of Ship Figure", IEEJ Transactions on Electronics, Information and Systems, Vol.140(3), pp.391-397, 2020.

夏 有輝也, 濱上 知樹, 菅家 正康, 吉田 健人, 庭川 誠, "車速追従制御のための強化学習における転移可能な方策の学習手法", 電気学会論文誌 C Vol.141(12) pp.1492-1499, 2021

5. 主な発表論文等

〔雑誌論文〕 計4件（うち査読付論文 3件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 兵頭 幸起, 濱上 知樹	4. 巻 141
2. 論文標題 ケアプラン作成支援システムのための非負値行列因子分解に基づく特徴語補完	5. 発行年 2021年
3. 雑誌名 情報処理学会論文誌	6. 最初と最後の頁 369-377
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 無
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 Sasaki Hayato, Yamamoto Mizuki, Takeshima Teppei, Yumura Yasushi, Hamagami Tomoki	4. 巻 139
2. 論文標題 Sperm Detection with Robustness to Overlap of Distributions by using Adaptive Thresholded Boosting	5. 発行年 2019年
3. 雑誌名 IEEJ Transactions on Electronics, Information and Systems	6. 最初と最後の頁 1461 ~ 1467
掲載論文のDOI (デジタルオブジェクト識別子) 10.1541/ieejeiss.139.1461	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 Lee Kundo, Hamagami Tomoki	4. 巻 140
2. 論文標題 Block-Based Neural Network Optimization with Manageable Problem Space	5. 発行年 2020年
3. 雑誌名 IEEJ Transactions on Electronics, Information and Systems	6. 最初と最後の頁 68 ~ 74
掲載論文のDOI (デジタルオブジェクト識別子) 10.1541/ieejeiss.140.68	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 Li Xin, Arai Hiroshi, Hamagami Tomoki	4. 巻 140
2. 論文標題 Wave-making Resistance Estimation Through Deep Learning Considering the Distribution of Ship Figure	5. 発行年 2020年
3. 雑誌名 IEEJ Transactions on Electronics, Information and Systems	6. 最初と最後の頁 391 ~ 397
掲載論文のDOI (デジタルオブジェクト識別子) 10.1541/ieejeiss.140.391	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 -

〔学会発表〕 計9件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 夏有輝也, 濱上知樹, 菅家正康, 吉田健人, 庭川誠
2. 発表標題 階層型強化学習と方策蒸留を用いた車両速度追従制御
3. 学会等名 令和3年電気学会全国大会
4. 発表年 2021年

1. 発表者名 Koichi Iwai, Masanori Akiyoshi, Tomoki Hamagami
2. 発表標題 Structured Feature Derivation for Transfer Learning on Credit Scoring
3. 学会等名 IEEE SMC2020 (国際学会)
4. 発表年 2020年

1. 発表者名 五井野琢也, 濱上 知樹
2. 発表標題 BERTを用いた医療文書からの固有表現抽出
3. 学会等名 第48回知能システムシンポジウム
4. 発表年 2021年

1. 発表者名 木村慶豪, 濱上 知樹
2. 発表標題 ファジィ測度を使用したアクティブラーニングによる物体識別
3. 学会等名 第48回知能システムシンポジウム
4. 発表年 2021年

1. 発表者名 山下廣大, 濱上知樹
2. 発表標題 エキスパート群から方策推定を行う敵対的逆強化学習
3. 学会等名 電気学会システム研究会
4. 発表年 2020年

1. 発表者名 藤井仁花, 濱上知樹
2. 発表標題 帯域別 pix2pix を用いた眼底異常検出
3. 学会等名 第29回インテリジェント・システム・シンポジウム
4. 発表年 2019年

1. 発表者名 板橋雅弘, 寺尾勇一, 林慧子, 佐野貴洋, 竹林奈々子, 濱上知樹
2. 発表標題 意味情報に基づく小規模ベイジアンネットワークを用いたケアマネジメントの評価値予測
3. 学会等名 第29回インテリジェント・システム・シンポジウム
4. 発表年 2019年

1. 発表者名 登内雅人, 新井洋, 濱上知樹
2. 発表標題 CNN を用いた類似した三次元船型形状からの造波抵抗推定
3. 学会等名 第18回情報科学技術フォーラム
4. 発表年 2019年

1. 発表者名 佐々木勇人, 岸大輝, 山本みずき, 竹島徹平, 湯村寧, 濱上知樹
2. 発表標題 ランキング学習を用いた顕微授精時の精子選別尺度獲得の検討
3. 学会等名 第18回情報科学技術フォーラム
4. 発表年 2019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

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