

## 科学研究費助成事業 研究成果報告書

令和元年6月3日現在

機関番号：14603

研究種目：若手研究(A)

研究期間：2016～2018

課題番号：16H05876

研究課題名(和文) 物理接触する人支援ロボットのための支援戦略モデルベース強化学習

研究課題名(英文) Model-based Reinforcement Learning of Assistive Strategies for Physically-assistive Robots

研究代表者

松原 崇充 (MATSUBARA, Takamitsu)

奈良先端科学技術大学院大学・研究推進機構・特任准教授

研究者番号：20508056

交付決定額(研究期間全体)：(直接経費) 14,300,000円

研究成果の概要(和文)：本研究では、物理接触を通じて人を支援するロボットが、被支援者に快適・安全な支援ができるよう、数回程度の実験試行から得られる非常に少ないデータによって、ロボットの制御規則(支援戦略)の自動設計を可能にする技術の確立を目指した。「少量データから実際上の最適戦略の学習」を目指すというパラダイムシフトにより、実用性の高い革新的技術の開発を試みた。その成果として、サンプル効率が高く、物理接触する人支援ロボットの支援戦略学習に適した強化学習アルゴリズムを開発した。さらに、安全性の高い膝動作支援デバイスを開発し、これを用いた支援戦略学習の被験者実験を通じて、提案手法の有効性を確認した。

研究成果の学術的意義や社会的意義

本研究では、人支援ロボットの支援戦略学習問題に対して、新しい思想に基づく技術開発を目指した結果、独創性と実用性を兼ね備えた強化学習アルゴリズムの開発に成功した。さらに、評価実験を実施するにあたり開発した膝関節支援デバイス自体も、モータの反トルクによる安全性の高い支援や、CVTの減速比調節によるユーザや状況への適応可能性など、高い学術的価値を持つ。本研究で得られた成果は、様々な人物理支援ロボットの支援戦略を被支援者にテイラーメイドに構築できる可能性を示唆しており、超高齢社会に有益な新しい産業・サービスの創出にも繋がるものと期待される。

研究成果の概要(英文)：In this research, we aimed to develop model-based reinforcement learning of assistive strategies for physically-assistive robots with data obtained through physical human-robot interactions. With the key concept of "Learning a practical (sub)optimal strategy from a small amount of data", we developed sample-efficient and model-based reinforcement learning algorithms based on Gaussian processes. Then, we developed a safe-assistive device for a knee joint and conducted learning experiments of assistive strategies for physically-assistive robots with human subjects. The experimental results demonstrated the effectiveness of the proposed approach.

研究分野：ロボットラーニング

キーワード：支援ロボット 強化学習 ガウス過程 ベイズ最適化 エナジーハーベスター

## 1. 研究開始当初の背景

超高齢社会を迎え、物理的な接触を通じて人を支援する「物理支援ロボット」に期待が高まっている。着用型・搭乗型・移動作業型などの形態があり、歩行や起立運動などタスクや用途も多岐に渡る。物理支援ロボットには「状況に応じて適切な支援行動を選択する支援戦略（ロボット制御規則）」が必要となる。しかし、適切な支援戦略は被支援者の状態、体格や感性に強く依存するため、その設計には被支援者と専門技術者による長時間の調節実験が不可欠である。これが物理支援ロボットの普及を妨げる大きな要因となっている。よって、支援戦略の自動設計は、実用化に向けて達成すべき重要な技術課題と位置づけられる。

被支援者-ロボット間の物理接触インタラクションのモデル化は複雑であり、モデルの要らない強化学習のアプローチが有望だと考えた。一般的な強化学習は「大量データから理論上の最適戦略の学習」を方針とするため、データ収集が被支援者に相当な負担を与える物理支援ロボットには適さない。そこで従来の枠組みに捕らわれず、例えば「数回程度の施行から得られる非常に少ないデータから、快適性と安全性を十分に満たす「少量データから実際上の最適戦略の学習」を目指すほうが実用性の高い技術になり得るとの着想に至った。

## 2. 研究の目的

本研究の目的は、物理接触を通じて人を支援するロボットが、被支援者に快適・安全な支援ができるよう、数回程度の実験試行から得られる非常に少ないデータによって、ロボットの制御規則（支援戦略）の自動設計を可能にする技術の確立である。データから制御規則の自動設計を行う「強化学習」に基づく従来技術では「大量データから理論上の最適戦略の学習」を方針とするため、データ収集が被支援者に相当な負担となる物理支援ロボットには適さない。そこで本研究では「少量データから実際上の最適戦略の学習」というパラダイムシフトにより、実装可能性の高い革新的技術の開発を目指す。

## 3. 研究の方法

主に次の二つの課題解決に取り組んだ。

### (1) データ不足を考慮するモデルベース強化学習の理論構築

わずかなデータから実際上の最適戦略を学習する鍵として「データ不足性」を考慮する新しい強化学習理論を構築する。予めデータ量が不十分とわかっている場合に、データ不足領域を戦略学習の対象範囲から除外する。これにより、理論上の最適戦略の学習に失敗するリスクを低減し、データ量に適った支援戦略を確実に学習するアイデアである。種々の強化学習の中で特に学習効率の良いモデルベース強化学習に着目する。データからシステムの状態遷移モデル（以下、モデル）を学習し、その後戦略を学習する2段構成である。データ不足を考慮するモデルベース強化学習理論の構築、計算効率の高いアルゴリズムの導出、最後にシミュレーション実験を経て、理論の完成を目指す。

### (2) 快適性・安全性を満たす支援戦略の自動設計技術の確立と実機検証

上記の強化学習理論に基づき、快適性と安全性を満たす支援戦略をわずかなインタラクションデータから自動設計する技術を確立する。快適性を満たすために、被支援者が与える評価値、もしくは、筋電位などの生体信号を強化学習の報酬とすることで実現する方法を開発する。センサ構成やどのような情報をユーザからの評価値として計測・フィードバックするかなどの方式も確立する。次に安全性を満たすために、ロボットが試行錯誤的に支援行動を実行してデータ収集する際と、学習された戦略の適用時に、被支援者に過度な負荷や無理な支援が行われる危険を低減する方法を検討する。人を物理的に支援するロボットシステムに加えて、支援技術の評価システムを統合した実験プラットフォームを構築する。被験者実験により提案手法の有効性や潜在する問題点を定量的に明らかにする。

## 4. 研究成果

### (1)に関する成果：

わずかなデータからモデルを学習する際に「データ不足性」を陽にモデル化するために、確率分布を利用する枠組みを検討した。システムの状態遷移をガウス分布と仮定すると、ガウス過程回帰と呼ばれる機械学習技術が利用できる。データの充足領域はモデルの予測分散が低く、不足領域は高く推定されるため、「データ不足性 = 分散」とした合理的にモデル化できる。まず、ガウス過程回帰によって制御対象のモデルを学習し、強化学習に利用するモデルベース強化学習の枠組みを検討した。さらに、EM 方策探索と呼ばれる方策ベースの強化学習手法に実装することで、従来手法よりも少ないサンプル数でも効率的に制御方策が学習可能であることを物理シミュレーションで確認した。次に、物理接触を含む制御問題を扱うために、非線形方策を学

習する枠組みについて検討した。具体的には、EM 方策探索の枠組みに基づいて、ガウス過程帰帰を方策モデルとして利用できるように拡張した。物理接触を通じて人を支援するロボットの支援戦略の強化学習に適した、高いサンプル効率と複雑な非線形方策の取り扱える枠組みが構築できたと考えられる。

2 年目以降では、ガウス過程方策探索の改良を実施した。先述の手法では、方策モデルとしてガウス過程を用いていたが、人の多様な行動を支援する方策としては表現能力が不十分であることが予備実験を通じてわかってきた。特に、同一状態において、最適な支援行動に複数の候補がある場合、ガウス過程ではそれらの平均を捉えてしまい、結果的に不適切な支援行動を学習する問題を確認した。その問題への対処として、重複混合ガウス過程を方策モデルとして利用し、その方策改善を実行できる強化学習アルゴリズムを開発した(図 1 参照)。また、別の問題として、筋電やモーションキャプチャなどのセンサデータから方策の評価信号である報酬値を算出する際には、ノイズや外乱の問題が避けられないこともわかってきた。この問題に対して、尤度関数にステューデントの  $t$  分布を用いるガウス過程方策探索のアルゴリズムを導出した。数値シミュレーションにより、その有効性および報酬ノイズへの頑健性を確認した。

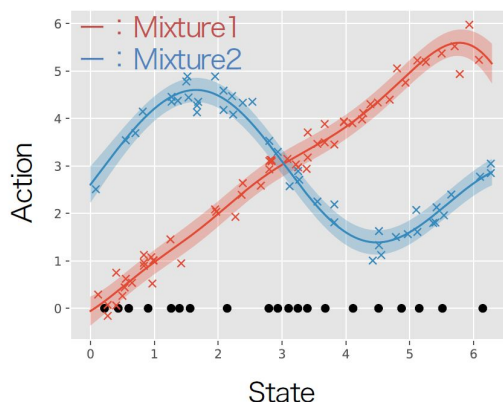


図 1 重複混合ガウス過程による方策探索

(2)に関する成果：

支援戦略の学習実験用のプラットフォームとして、安全性に配慮して、モーターの回転抵抗を制動力として利用する回生ブレーキによる膝関節支援装置型ロボットを設計し、そのプロトタイプを開発した(図 2 参照)。被支援者の受ける負荷や筋電図を同時計測するように、センサおよび計測システムの統合も行った。本質的に安全性が高いため、支援戦略の学習技術の有用性を被験者実験によって実施する際に、有用なシステムが構築できたと考えている。

次に、開発した型膝動作支援デバイスを用いて、被験者実験を実施し、提案技術の有効性や安全性等を定量的に評価した。支援方策の評価については、被験者の筋電量およびロボットの回生エネルギーに着目した基準を考案した。支援戦略の学習アルゴリズムとして、(1)にて開発したアルゴリズムの簡略版として、ガウス過程に用いたベイズ最適化によって、パラメトリックな関数で設計された方策のパラメータを、安全性および支援効果を最大化するように学習した。複数被験者ならびに様々な路面状況において、比較的少量の実験データのみから、ユーザや状況に応じたテーラードな支援が学習できることを実験的に確認した。

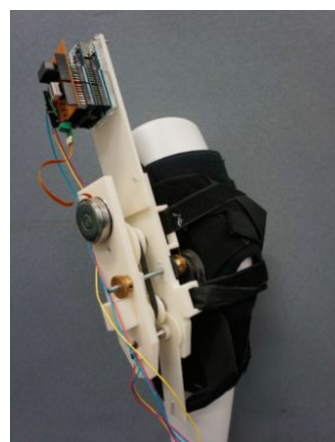


図 2 開発した膝動作支援デバイス

## 5 . 主な発表論文等

〔雑誌論文〕(計 2 件)

Takamitsu Matsubara, Kotaro Shibata: Active tactile exploration with uncertainty and travel cost for fast shape estimation of unknown objects. *Robotics and Autonomous Systems* 91: 314-326 (2017) 査読有り

Ming Ding, Takamitsu Matsubara, Yoshihito Funaki, Ryojun Ikeura, Toshiharu Mukai, Tsukasa Ogasawara: Generation of comfortable lifting motion for a human transfer assistant robot. *IJIRA* 1(1): 74-85 (2017) 査読有り

〔学会発表〕(計 8 件)

佐々木光, 松原崇充,  
“ スチューデントの  $t$  分布を尤度関数に用いたロバストなガウス過程方策探索, ”  
平成 30 年度 SICE 関西支部・ISCIE 若手研究発表会, C3-1, pp. 166-170 (USB), 2019

Hikari Sasaki and Takamitsu Matsubara: Multimodal Policy Search using Overlapping Mixtures of Sparse Gaussian Process Prior, IEEE ICRA 2019, pp. 2433-2439.

Yutaro Ikawa, Taisuke Kobayashi, Takamitsu Matsubara: Biomechanical Energy Harvester with Continuously Variable Transmission: Prototyping and Preliminary Evaluation, IEEE AIM 2018, pp. 1045-1050

Takamitsu Matsubara, Yu Norinaga, Yuto Ozawa, Yunduan Cui: Policy Transfer from Simulations to Real World by Transfer Component Analysis, IEEE CASE 2018, pp. 264-269.

Hikari Sasaki, Yuto Ozawa, Takamitsu Matsubara: Variational Learning Approach for Sparse Gaussian Process Policy Search, SICE Annual Conference 2018, pp. 642-645.

佐々木光, 小澤裕斗, 松原崇充,  
“変分学習によるスパース擬似入力ガウス過程方策探索”  
ロボティクス・メカトロニクス 講演会 2018, 1A1-C16, 2018

井川優太郎, 小林泰介, 松原崇充,  
“CVT 付装具型エネルギーハーベスタにおける変速比のベイズ最適化”  
第 36 回日本ロボット学会学術講演会, 1C2-02, 2018

佐々木光, 松原崇充,  
“重複混合スパースガウス過程を事前分布とする多峰性方策探索”  
第 36 回日本ロボット学会学術講演会, 1C3-03, 2018

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況(計 0 件)

## 6. 研究組織

なし

科研費による研究は、研究者の自覚と責任において実施するものです。そのため、研究の実施や研究成果の公表等については、国の要請等に基づくものではなく、その研究成果に関する見解や責任は、研究者個人に帰属されます。