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　４ 年　 ６ 月　 ２ 日現在

機関番号：１３９０１
研究種目：若手研究
研究期間：2019～2021
課題番号：１９Ｋ２０３７５

研究課題名（和文）統計的機械学習の手法を用いたデータ駆動型非線形準最適制御

研究課題名（英文）Data-driven quasi-optimal control using machine learning techniques

研究代表者
有泉　亮（Ariizumi, Ryo）

名古屋大学・工学研究科・助教

研究者番号：３０７７５１４３

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）：ロボットなどへの応用を念頭に，比較的少ない実験回数で最適な制御入力を得る強化学習則を目指し研究を行った．特に，ロボットの強化学習法として知られるPI2と呼ばれる強化学習則の応用を中心に検討した．これにより，従来の強化学習法では学習困難であった脚ロボットの転倒状態からの起き上がり動作の習得など，困難なタスクを数千回程度の実験結果をもとに達成することに成功している．また，制御工学の知見を応用することにより，より効率よく学習を行うための基礎的な検討を行った．

研究成果の学術的意義や社会的意義
強化学習の有効性は様々な分野で明らかになってきているが，多自由度ロボットの強化学習は状態や入力が連続値であることもあり，タスクによっては数十万回に及ぶ実験が必要となるなど，まだ実用に足る効率は発揮できていない．本研究ではデータ効率の向上を目的に，データの使い方の工夫を提案した．また，データの工夫だけでは効率化に限界がある．そこで，明らかに成立する物理的性質を学習に取り入れることを考え，その実現のための基礎的検討を行った．これらは，今後さらに強化学習の効率を向上させ，多自由度ロボットの強化学習のデータ効率を実用的なレベルに引き上げるための基礎となりうる．

研究成果の概要（英文）：In this research, we aimed to propose reinforce learning methods that can obtain sub-optimal inputs (actions) with a relatively small number of samples. Especially, we put our attention on the PI2 algorithm, which is known to be efficient for robots with large degrees of freedom. One of our proposed algorithms achieves a standing-up motion of a legged robot, which is turned over at the initial state. This task is very difficult for most existing methods, but our method succeeded by using a few thousand samples. We also conduct a basic study to employ control-theoretic methods to speed-up reinforcement learning.

研究分野：ロボティクス

キーワード：強化学習　ロボティクス　制御工学

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景
移動ロボットの運動最適化など複雑なシステムの最適制御問題の多くは，解析的な手法の適用が困難である．このため，強化学習の応用が広く考えられている．しかし，既存の強化学習では膨大なデータを必要とし実用的とは言い難い．そのため，比較的少ないサンプルから最適化を実行する手法が望まれている．我々のグループではこの問題に対し，主に応答曲面法（ベイズ最適化）による方法を提案してきていたが，扱えるパラメータ数が少なく，ロボットの運動最適化において十分とは言い難い状況である．

２．研究の目的
比較的少ないサンプルから最適制御入力を得る，データ駆動型準最適制御法を提案する．またその実現のために，システムの簡略化モデルなど，容易に入手可能な事前情報を適切に利用する方法を提案する．

３．研究の方法
目的達成のために，でータの扱い方の効率化，データ取得法の効率化，システムモデルの適切な応用と，データ駆動モデル構築の 3 点から考察を進めた．上記のうち，最初 2 点については，両方を同時に達成する強化学習則の提案を進め，3 点目についてはデータ駆動モデル構築について重点的に考察を進めた．なお，強化学習アルゴリズムのベースとしては，代表者が今まで研究に携わってきた応答曲面法と PI2（Policy Improvement with Path Integral）に着目した．

４．研究成果
大きく，(1) 高自由度ロボット用の強化学習アルゴリズムとして知られる PI2 のさらなる高効率化の達成，(2) 物理システムの構造的な特徴を強化学習に利用するための基礎的検討およびその有効性確認，(3) 高速データ駆動システム推定に関する理論的検証，という三つの成果を得ている．特に，PI2 の高効率化に関しては，他の既存の強化学習法では達成困難な学習に成功している．その成果は学習分野のトップクラスのジャーナルに採択されている．なお，研究開始当初に最も有力と考えていた応答曲面法の応用に関しては，PI2 への応用や深層学習との融合など様々な方向も模索したが，芳しい成果は得られなかった．これは，システムの簡略化モデルまでもデータ駆動で作成することを想定していたが，その方法の確立が難しいという点に最大の問題がある．現在，成果 (2) に関連して考案中の方法や (3) で考察の対象とした方法を発展させることで，この問題を解決できないかと考え研究を継続中である．他に，ニューラルネットワークの情報を圧縮するなどにより深層強化学習を効率化する方法についていくつか考えていたが，いずれも成功には至らなかった．以下では期間内に得られた成果 (1)，(2)．(3) について記述する．

(1) 強化学習アルゴリズム PI2 の更なる高効率化
PI2 は確率的最適制御を応用して導出された強化学習法で，特に生物模倣ロボットなどの高自由度のシステムに対して有効な方法として期待されており，脚ロボットの動作学習（Theodorou et al., Int. J. Machine Learn., 2010）や，ヘビ型ロボットの一種であるねじ推進ヘビ型ロボットの動作学習（Chatterjee et al., Int. J. Adv. Robot. Sys., 2015）などに応用されている．PI2 は強化学習法の中ではハイパーパラメータが比較的少ない手法ではあるが，3 種類のハイパーパラメータが存在し，それらいずれに対しても学習効率は敏感に変化する．このため，実際にこの学習法で適切な運動を習得することは必ずしも容易ではなかった．

PI2 では，ロボットの動きをパラメータ化し，そのパラメータの最適化を達成するという問題を扱う．このために，ある正規分布（探索分布）からのサンプルによっていくつかのパラメータを決定し，それらのパラメータについて実験を行い，その実験結果の評価値を用いて探索分布を更新する，という手順を繰り返す．PI2 に存在する 3 種類のハイパーパラメータとは，探索計画に用いる探索分布の共分散行列，アップデートに使うサンプルの個数（世代あたりの個体数），取得したデータの内部での扱われ方を決める値（温度パラメータ）である．これらのうち，共分散行列については Stulp ら（Stulp and Sigaud, ICML, 2012）が，データに合わせて自動調整する手法 PI2-CMA を提案している．この研究では PI2 と進化戦略との類似性に着目して，進化戦略の一種である CMAES の更新方法を応用している．

本研究では，PI2-CMA をさらに発展させ，世代あたりの個体数も自動調整する手法を提案した．この方法では個体数調整だけではなく，軌道にそったステージコスト系列の類似性などを利用し，比較的スパースなデータを利用した処理を行うなどの工夫も加えることで，データの利用効率の高さと計算コストの低さを両立させることに成功している．また，Stulp ら（Stulp and Sigaud, ICML, 2012）の PI2-CMA における共分散行列更新の方法をさらに発展させ，よりきめ

細かな調整が行えるように改良を施した．この方法では，従来の強化学習法では獲得困難であった，脚ロボットの転倒状態からの起き上がり動作の学習を数千回のデータで完了させるなど，高いパフォーマンスを確認してる（図 1）．この結果は IEEE Trans．Cyb．に掲載済みである．一方，この方法を用いても，実機実験による学習は現実的ではなく，シミュレーションによる学習が必要である．しかし，シミュレーションと実機では一般には挙動が異なり，シミュレーションを基に得られた学習結果が必ずしも実機に対して適切とは限らない．そこで，モデル化誤差に対してロバストな動作を学習する手法も提案している．この方法については，環境との摩擦の影響に対し動作が敏感に変化するロボットを利用して，有効性を確認している．この結果はシステム制御情報学会誌にて発表済みである．

図 1　提案法で獲得した 5 脚ロボットの転倒状態からの起き上がり動作

さらに，PI2 の残る一つのハイパーパラメータである温度パラメータについても自動調整する方法を提案している．この方法は，PI2 の導出にまで立ち返り，PI2 のアルゴリズム上に含まれていた矛盾点を解消する試みから発想を得たものであり，それ単独で非常に大きな学習効率改善を確認している．結果の一部は国内会議 SCI ’ 21 にて発表している（指導学生が発表し，システム制御情報学会研究発表講演会学生発表賞を受賞）．また，成果をまとめたものは学術雑誌の査読を受け，修正中である．

(2) 物理システムの構造的特徴を利用した強化学習に関する基礎検討
シミュレーションを基にロボットの動作獲得を実行した場合，得られた動作が実機に対しても適切なものである保証はない．これは，モデル化誤差が原因であり，モデル化誤差によってシミュレーションに固有に発生する現象をも利用した挙動を学習してしまうことに問題がある．この点は良く知られており，解決のために様々な研究がなされているが，そのほとんどは，幅広いモデルでの学習を行う，あるいは，シミュレーションが信頼できる範囲と信頼できない範囲を明確にすることで，実機における不適切な挙動を防ぐというものである．しかし，代表者はこの問題がモデルの細部を使いすぎていることに起因する，と言えるのではないかと考え，支配方程式の構造などより大まかな性質を利用できれば解決できるのではないかと考えている．

そこで，本研究では受動性と呼ばれる，保存則の拡張ともいえる性質に着目し，ポート・ハミルトン形式（PH 形式）と呼ばれるシステムモデルを利用することを考えた．受動性はほとんどの物理システムが自然に満たす性質であり，投入したエネルギー以上のエネルギーを取り出すことができない，ということに対応している．一方，この性質を表現するために便利なシステム表現として，ハミルトンの運動方程式の拡張である PH 形式がある．PH 形式を活用した制御則は，ロバスト制御の分野で活発に議論されている．PH 形式の構造的特徴を利用した制御器を作成し，そのパラメータを強化学習により調整する，とすることで，モデルの細部が実機と異なってても，実機で問題なく使用できる挙動が得られると期待できる．また，シミュレーションによる動作獲得のみを考えても，モデルを利用することで学習効率の向上が期待できる．

本期間では，まず PH 形式を応用したヘビ型ロボットの制御則について考案した．PH 形式に基づく制御則は広く研究されているが，ヘビ型ロボットのような生物模倣ロボットの多くは，従来の研究で想定されている条件を満たさず，そのままでは適用できない．そこで本研究では，従来知られていた方法に対し，システムの支配方程式が持つ構造の違いから生じる問題を解消する拡張を提案した．また，シミュレーションによりその有効性を確認した．この結果は AROB Journal にて発表済みである．また，この方法はモデル化誤差に対してロバストであることが期待され，この点が強化学習に活用できると考えている．ロバスト性については，その成立を示唆するシミュレーション結果が得られている（上記 AROB Journal に掲載済み）ほか，理論的に定量評価を試みた．ロバスト性に関する理論検証に関しては，現在，細部を修正し論文誌への投稿を目指している．

一方，本来の目的である強化学習への応用に関してはまだ，2 種類のシミュレーションを利用した予備的な検証を実施した段階であるが，期待していたような効果が実際に得られることを示唆する結果が得られている．今後，実機検証を実施するなど，より説得力のあるデータを集める．

(3) 高速データ駆動システム推定に関する理論的検証
近年，動的システムの高速学習モデルとして，リザバーコンピューティング （RC） が期待され

ている．これは，リザバーと呼ばれる動的システムとリードアウトと呼ばれる静的システムを組み合わせた学習モデルであり，リザバーは学習せずリードアウトのみをデータに合わせて調節する，という点に特徴がある．動的システムであるリザバーに対しては訓練を行わないため，学習は単純な回帰問題や分類問題に帰着でき，小さな計算コストで学習できる．しかし，このようなシステムで本当に学習が可能なのか，という点が大きな問題である．これに対し，Maass ら (Maass et al., Neural Comput., 2002) は，近似対象のシステムがフェーディングメモリという特性を持つとき，リザバーもフェーディングメモリを有し，かつ，リードアウトが分離特性と呼ばれる特性を有していれば，任意精度で近似対象を近似できる（計算万能性を有する）ことを示した．しかし，リザバーがフェーディングメモリを有するというのはかなり強い拘束であり，もしその条件が不要であることを示すことができれば，RC の活用範囲を広げることができる．

本研究では，リザバーが明らかにフェーディングメモリ特性を持たないにも関わらず，RC が適切に適用できる例を示した．また，近傍分離特性という新たな概念を提唱し，リードアウトが近傍分離特性を満たしていれば，リザバーのフェーディングメモリ特性がなくとも計算万能性を有することを数学的に証明した．このために，関数解析の有名な定理である Stone-Weierstrass の定理を拡張した新しい関数近似定理を証明した．加えて，近傍分離特性は Maass らの提示した条件より真に緩い条件であることを示した．これにより，今までに試みられている多くの RC モデル（多くの場合，次段落で説明する理由により Maass らの条件を満たさないと推察している）は実際に適切な近似器になっている可能性が示唆される．この内容の一部は国際会議 ASCC2022 にて発表済みである．また，証明の細部など，全体をまとめたものは国際学術雑誌に投稿中である．

さらに，上記に加え，Maass らの提示した条件は有限次元の RC モデルでは実現不可能であること，本研究で新たに示した条件であれば実現不可能とは言えないこと，もし本研究の条件を満たすならばリザバーはカオス的挙動に近くなること，の 3 点を証明できると考えている．現在，RC を応用する多くの論文では，RC を適用する根拠として Maass らの条件を挙げたうえで，その条件の成否については全く検証しないまま議論を進めているか，RC を適用できるとする根拠について何一つ議論していないかのいずれかである．また，Maass らの条件を満たす，という方向からは矛盾するようなカオス的システムに近いリザバーの方が，適切な学習結果を得やすいことを示唆する結果も多く発表されている．本研究の証明が完成すれば，このような研究結果に対して，理論的に妥当な解釈を与えることができる．この内容に関しては現在，証明の大枠は完成したと考えているが，細部の検討を進めている．

５．主な発表論文等

〔雑誌論文〕　計2件（うち査読付論文　2件／うち国際共著　0件／うちオープンアクセス　1件）

１．著者名 Yamamoto Kosuke、Ariizumi Ryo、Hayakawa Tomohiro、Matsuno Fumitoshi	４．巻 52
２．論文標題 Path Integral Policy Improvement With Population Adaptation	５．発行年 2022年
３．雑誌名 IEEE Transactions on Cybernetics	６．最初と最後の頁 312-322
掲載論文のDOI（デジタルオブジェクト識別子） 10.1109/TCYB.2020.2983923	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 藤原大悟，山本耕輔，有泉亮，早川智洋，松野文俊	４．巻 33
２．論文標題 モデル化誤差に対してロバストな学習のためのコスト関数更新手法の提案	５．発行年 2020年
３．雑誌名 システム制御情報学会誌	６．最初と最後の頁 191-200
掲載論文のDOI（デジタルオブジェクト識別子） 10.5687/iscie.33.191	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3件（うち招待講演　0件／うち国際学会　2件）

１．発表者名 Yasuhiro Imagawa, Ryo Ariizumi, Toru Asai, Shun-ichi Azuma
２．発表標題 Port-Controlled Hamiltonian Approach for Robust Control of Snake Robots
３．学会等名 The 4th International Symposium on Swarm Behavior and Bio-Inspired Robotics（国際学会）
４．発表年 2021年

１．発表者名 中野裕康，有泉亮，浅井徹，東俊一
２．発表標題 経路積分に基づく直接方策改善法による強化学習における温度パラメータの自動調整
３．学会等名 システム制御情報学会研究発表講演会
４．発表年 2021年

１．発表者名 Shuhei Sugiura, Ryo Ariizumi, Toru Asai, Shun-ichi Azuma
２．発表標題 Universality of reservoir computing using reservoirs without fading memory
３．学会等名 Asian Control Conference（国際学会）
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

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