

令和元年6月12日現在

機関番号：13901

研究種目：若手研究(B)

研究期間：2017～2018

課題番号：17K14622

研究課題名(和文) 運動学習プリミティブと応答曲面法による冗長ロボットの最適運動決定

研究課題名(英文) Movement Optimization of Redundant Robots using Dynamic Movement Primitives and Response Surface Method

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) 3,400,000円

研究成果の概要(和文)：本研究では応答曲面法を用いた冗長ロボットの強化学習法を確立することを目的とした。応答曲面法は少ないサンプルを基に最適パラメータを探索でき、特に低次元問題に有効である。冗長ロボットの場合、時間などの問題から多数のサンプルを得ることは難しいことから、応答曲面法の適用が望ましい。ところが、一般に決定変数空間が高次元であり、応答曲面法の適用は容易ではない。そこで、低計算コストでありサンプル回数に制約は無いが信頼性の低いシミュレーションと、信頼性は高いがサンプル数が限られる高精度なシミュレーション・実験とを併用して、適切に学習を進める方法を提案した。

研究成果の学術的意義や社会的意義

ロボットの運動決定法として強化学習が注目されているが、多数のサンプルを必要とし、実際のロボットへの適用は難しい。このため、簡易なシミュレーションを活用することが考えられるが、モデル化誤差が実システムでの学習に悪影響を及ぼすことが知られている。本研究で提案する手法では、モデル化誤差による悪影響を回避しながら、簡易シミュレーションを有効に活用して、実システムにおけるサンプル回数の大幅な削減を実現する。これにより、既存の強化学習では適用困難であった高次元なロボットシステムに対しても、強化学習による最適運動の探索が可能となる。

研究成果の概要(英文)：In this research, a reinforcement learning method for a redundant robot based on the response surface method (RSM) is proposed. RSM is known to be a data efficient optimization method for relatively low dimensional systems. For redundant robots, because it is difficult to take many samples, RSM is expected to be suitable. However, high dimensionality of the problem prevents RSM from being applied to such robots. To solve this problem, we proposed to use low-cost simulations, whose number of samples is not restricted, along with the actual simulations/experiments, appropriately to prevent the negative effects of the error in the low-cost simulations.

研究分野：知能ロボティクス

キーワード：応答曲面法 ロボット 機械学習

様式 C-19、F-19-1、Z-19、CK-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、評価関数の数理的表現に基づく最適化が困難な問題に対し、強化学習の応用が考えられている。シミュレーションではモデル化誤差が生じることから、実験のデータを基に最適解を探索できることが望ましいが、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、実機実験の必要回数をシミュレーションの利用により削減しつつも、シミュレーションの持つモデル化誤差の影響を抑える手法が望まれている。

2. 研究の目的

データ駆動の最適化法、特に応答曲面法とそこで用いる事前知識の与え方に着目し、シミュレーションの持つ誤差の影響を抑えながら、シミュレーションと実機実験を組み合わせた最適探索法の提案を目指す。そのために、あるシステム A を近似する別のシステム B が存在したときに、B の観測結果を適切に利用しながら A に対する最適解を発見する手法を提案し、有効性を検証する。

3. 研究の方法

応答曲面法（ベイズ最適化とも呼ばれる）における応答曲面推定法に着目する。応答曲面法は観測データを基に評価関数を決定変数の関数（これを応答曲面と呼ぶ）として推定し、その推定を用いて次の観測の際に用いる決定変数を計画し観測を加える、という操作を繰り返す逐次的な最適化法である。応答曲面推定には任意の回帰法が使用できるが、ガウス過程回帰を用いるとする場合、事前分布を設定できる。この事前分布の構成に対象の近似システムによる観測結果を用いるとすることで、近似誤差の影響はあまり受けずに応答曲面推定の精度を向上できると予想される。一方で、実際のシステムと近似システムとの誤差があまりに大きいと、逆効果になる可能性も考えられる。この点についても考察を行う。

4. 研究成果

大きく、次の（ア）—（エ）の成果を得ている。いずれもまだ検討の余地が残っているが、（ア）、（イ）に関しては研究計画時からの研究の目的に沿った内容であり、その結果の一部をまとめたものは国際学会誌へ投稿中である。（ウ）、（エ）は当初の計画に沿った研究を進めていく中で出てきたアイデアに関するものである。

（ア）近似的なシステムを利用する方法として、応答曲面法作成の際に用いるガウス過程回帰の事前分布に着目した。ガウス過程回帰はベイズ推定の一種であり、これを用いることで未観測の決定変数に対する評価関数値の予測をガウス分布の形で与える。また、他のベイズ推定の手法と同様、事前分布の形で事前知識を反映させることも可能である。なお、ガウス過程回帰における事前分布としてはガウス過程を用いる。ガウス過程は平均関数と分散関数を定めることで指定される。応答曲面法においては、一般にシステムについての事前知識を仮定しないため、この事前分布としては平均関数が線形関数などの非常に簡潔な形で表現されているもののみを扱っていた。しかし、システムの応答の近似が容易に得られる場合には、その応答自体を平均関数に設定することで、事前知識を適切に扱うことが可能であると考えられる。

この方法を適用する際の問題点として、システムの近似精度がどの程度であれば十分なのか分からないという点が挙げられる。本研究では、まずその点の解明を目標として検討を行った。ガウス過程回帰の特性から、例えば近似システムを用いて計算した場合の評価値が、実システムを用いて計算した評価値に比べ、全体的に一定程度過大・過小評価される場合（評価関数に評価値方向の平行移動が生じる場合）、近似誤差の影響はほとんど存在しないことが推測され、この点については数値検証においても確認されている。

また、評価関数の停留点の位置に関する誤差（評価関数に決定変数空間方向の平行移動が生じる場合）についても、大きな影響は生じないと考えられる。この点についても数値検証では確認済みである。

一方で、評価関数値の「変化の激しさ」に関する誤差は大きく作用すると考えられる。この変化の激しさの指標に関しては、微分値などをその指標として使用できると考えられる。実際、決定変数が 1 変数の場合には、微分値誤差の最大値によって、事前分布が与える影響をある程度定量的に評価できることを数値例により確認している。これは、ガウス過程回帰による推定に関し、大胆な近似を導入して得られた結果から推論されたものである。しかし、決定変数が 2 変数以上の場合には類似の関係性は存在しないことが数値例から予想されている。1 変数の際に行った推論の大部分は決定変数の空間の次元には依存しないものであるが、推論の際に用いた近似が 2 変数以上になると十分な近似ではなくなるということであると推測される。この点については、研究期間終了まで検討を続けてきたが、まだ決定的な原因を発見するには至っていない。

以上の内容に関して、その考察の一部については次の項目（イ）で述べる内容と合わせて論文誌へ投稿中であるが、特に決定変数が 2 次元以上である場合の論証は完成していない。一方で、実際に応答曲面法を応用するシステムは一般には 2 次元以上であり、現状の議論では不十分である。このため、今後も検討を継続する必要がある。

(イ) 応答曲面法をロボットの運動最適化に適用する際、ほとんどの場合は簡単な運動のパラメータ化により 2 次元程度の問題に帰着させている。例えば、ヘビ型ロボットの運動最適化を扱った研究 [1] においては、サーペノイド曲線と呼ばれる、ヘビ型ロボットの体形曲線としてはデファクトスタンダードとなっている曲線を採用し、そのパラメータを調節するとした。しかしこのような方法では、あらかじめ適切な運動が分かっている場合は適用困難である。

そこで本研究では、より一般的な運動表現として運動学習プリミティブを採用した。運動学習プリミティブは、ロボットの強化学習の分野ではよく知られた運動のパラメータ化の一つである。その最適化には確率最適制御を応用した強化学習法である PI^2 [2] など、評価関数値の時系列変化を考慮した最適化法が一般に用いられてきている。ところが、純粋に問題の定式化を見れば単なる時不変パラメータの最適化であり、評価関数値の時系列を考慮する必要性は自明ではない。また、 PI^2 やその拡張において用いられている時系列の考慮法の妥当性について、理論的な裏付けは存在しない。一方で、時不変パラメータの最適化においてわざわざ時系列を考慮するために、評価関数の設定に工夫が必要となり、適用範囲を限定してしまう要因ともなる。本研究においては、そのような時系列の考察は行わずに、パラメータ最適化の手法である応答曲面法を適用することを提案した。

運動学習プリミティブの学習が単なるパラメータの最適化であること、ロボットの強化学習はデータ駆動の最適化であることに着目すれば、応答曲面法の利用は自然な考え方である。しかし、今までこのような組み合わせが存在していなかった理由として、運動学習プリミティブで十分幅広い運動を表現するためには、パラメータ数が多くなってしまふ点が挙げられる。この場合、応答曲面法の利点である、必要な実験回数の削減の効果を活かしにくくなってしまふ。

応答曲面法は低次元問題においては、強化学習や進化戦略などの他のデータ駆動型の最適化法と比べると、利用するデータ数ははるかに少なく済むが、それでも次元数の十倍から数十倍のデータが必要となる。例えば、ロボットの運動計画を 20 次元程度の決定変数空間における最適化問題に落とし込んだとすると、必要な実験回数は 200 回から 1000 回程度と予想される。それでも、一般的な強化学習（適切な初期解無しで適用する場合）と比べると大幅に少ないが、実機実験を基に最適化を行うには非現実的な回数となる。実機実験に基づく最適化が非現実的であり、実行回数に上限のない（あっても非常に大きい）シミュレーションに頼る場合、従来の強化学習法で十分ということになってしまう。このことから、問題の次元が高くなると、応答曲面法の利点は活かしにくくなる。

しかし、応答曲面法に用いるガウス過程回帰において、その事前分布の平均関数をシミュレーション結果に置き換えることで、必要な実験回数を大幅に削減可能であると考えた。また、 PI^2 などでは、シミュレーションを基に最適化を行った後、その解を初期値として実機実験を基にした強化学習を別途行うとすることが多い。しかし、このような方法では、シミュレーションに含まれるモデル化誤差の影響を受け、適切な解には収束しないことも多い。この問題はモデル化バイアスの問題として知られており、その解決策が活発に議論されているが、応答曲面法の事前分布にシミュレーション結果を用いるとする場合には、モデル化バイアスの影響は項目 (ア) の考察結果から、比較的小さいと予想できる。このことから、提案法は、強化学習におけるモデル化バイアスの軽減法としても重要であると考えている。

検証実験として、ロボットマニピュレータによるけん玉タスクの最適化を行った。研究計画の段階では譲り受けることになっていた PA-10 の実機を使用する予定であったが、急遽譲り受けることができなくなった。そこで、実機の代わりに物理シミュレータを使用するとし、物理シミュレータと比べ大幅に簡略化された運動学シミュレータを事前平均関数生成に使用するとした。事前平均関数として利用したシミュレータは、けん玉の球を取り損ねることが想定されていないなど、動力学シミュレータとは大幅な違いのある挙動をするものとした。運動学習プリミティブのパラメータ数を 20 として最適化を行い、 PI^2 などと比べ少ない実験（動力学シミュレータでの実行）回数で最適に近い運動を獲得できることを確認した。特に、けん玉タスクに成功するまでに要する実験回数は数十回程度であり、実機実験を用いた最適化も現実的と考え得る結果となった。獲得したけん玉タスクの様子を図 1 に示す。



図 1 獲得したけん玉タスクの様子

以上の内容については、一部を国内学会で発表したほか、まとめたものを国際学会誌へ投稿予定である。

(ウ) 本研究においては、応答曲面法の応用によるデータ駆動型最適化を考えており、応答曲面法の中でもベイズ最適化とよく呼称されるような手法を中心的に扱っている。ベイズ最適化はベイズ推定による回帰を用い、また、実験計画にベイズ推定の結果を用いているからであるが、実は応答曲面法に置いてベイズ推定の利用は必ずしも必須ではない。実際、ベイズ推定を一切利用せずに、同様の目標を達成する方法も提案されている [3]。ベイズ推定を必要としない枠組みを適切に構築できれば、回帰法はどのようなものでも構わず、例えば深層学習を応用することが可能である。特に、近年注目を集めている敵対的生成ネットワークにより応答曲面推定を行うとすることで、推定の不確かさも適切な形で評価できるのではないかと考えている。

この内容についてはもともと研究計画にはなかったものであるが、本格的に研究を進めていくための準備を進めているところである。準備の一環として深層学習の技術を用いた制御器の秘匿化を提案し、その結果は国際会議へ投稿中である。

(エ) 上記の項目で主に考察した内容は、ロボットのフィードフォワード型の最適制御を達成する方法であると考えることが可能である。しかし、実際にはセンサからの情報をもとにフィードバック制御を施す必要があると考えられる。そこで、データ駆動型最適フィードバック制御の構築を検討した。

このようなデータ駆動型最適フィードバック制御を応答曲面法を用いて行うためには、入力を適切にパラメータ化された関数のクラスに限定する必要がある。そのためのパラメータ化が最大の問題となる。項目 (イ) で使用した運動学習プリミティブのような方法を使用することになると考えており、状態が1次元であれば、運動学習プリミティブがそのまま使えると考えて、基礎検討を行っている。しかし、多次元の場合には拡張困難であり、現在、適切なパラメータ化に関する検討を進めている。

この内容に関してはもともと研究計画にはなかったものであり、上記のように非常に基礎的な検討にとどまっているが、今後の発展が見込めると考えている。

参考文献

- [1] R. Ariizumi, M. Tesch, K. Kato, H. Choset, and F. Matsuno: Multiobjective Optimization Based on Expensive Robotic Experiments under Heteroscedastic Noise, IEEE Transactions on Robotics, Vol. 33, No. 2, pp. 468 - 483 (2017)
- [2] E. A. Theodorou, J. Buchli, and S. Schaal: A Generalized Path Integral Control Approach to Reinforcement Learning, Journal of Machine Learning Research, vol. 11, pp. 3137-3181 (2010)
- [3] H.-M. Gutmann: A Radial Basis Function Method for Global Optimization, Journal of Global Optimization, Vol. 19, pp. 201-227 (2001)

5. 主な発表論文等

〔雑誌論文〕 (計 件)

〔学会発表〕 (計 1 件)

- ① 奥村, 有泉, 浅井, 東: 多関節ロボットによる柔軟動作の応答曲面法を用いた学習, 第5回制御部門マルチシンポジウム, 東京, 3月8~11日, Fr73-2 (2018)

〔図書〕 (計 件)

〔産業財産権〕

○出願状況 (計 件)

名称:

発明者:

権利者:

種類:

番号:

出願年:

国内外の別:

○取得状況 (計 件)

名称：
発明者：
権利者：
種類：
番号：
取得年：
国内外の別：

〔その他〕
ホームページ等

解説論文（計 1 件）

- ① 有泉：ガウス過程回帰を用いた多目的最適化とロボット制御への応用，システム/制御/情報，vol. 62, no. 10, pp. 429-434, 2018

招待講演（計 1 件）

- ① 有泉：ガウス過程回帰とロボットの運動パラメータ最適化への応用，第 63 回システム制御情報学会研究発表講演会（SCI' 19），中央電気倶楽部，5 月 24 日（2019）

6. 研究組織

(1) 研究分担者

研究分担者氏名：

ローマ字氏名：

所属研究機関名：

部局名：

職名：

研究者番号（8 桁）：

(2) 研究協力者

研究協力者氏名：

ローマ字氏名：

※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。