様　式　Ｃ－１９、Ｆ－１９、Ｚ－１９　（共通）

科学研究費助成事業　　研究成果報告書

平成 ２８ 年　 ６ 月　 ６ 日現在

機関番号：１２６０１
研究種目：若手研究(B)
研究期間：2013～2015
課題番号：２５７３０１３５

研究課題名（和文）データ同化強化学習

研究課題名（英文）Data assimilation based reinforcement learning

研究代表者
植野　剛（Ueno, Tsuyoshi）

東京大学・新領域創成科学研究科・特任研究員

研究者番号：９０６１５８２４

交付決定額（研究期間全体）：（直接経費）　　3,200,000円

研究成果の概要（和文）：計算機シミュレーションの結果から行動戦略を学習することは，多大なコストが発生する「実験」を行なう必要がないため，飛躍的な生産性の向上が見込まれる．しかし，シミュレーションの結果は実際の実験結果から乖離することも少なくなく，シミュレーションで得た戦略が実際に意味をなさないことも多い．本研究では，シミュレーション学習の枠組みであるデータ同化と，データから意思決定則を学習する強化学習を組み合わせた「データ同化強化学習」を開発し，少ない実験数で高い性能を発揮する行動戦略の学習を実現した．また，開発した手法を新規材料設計問題に応用し，従来法より少ない実験数で目標とする物質を見つけることに成功した．

研究成果の概要（英文）：Learning action strategies from computer simulations has a potential to achieving the drastic productivity increases because it has no necessity to perform an expensive process, i.e., real experiments for collecting the data. However, the behavior of simulations often differ from that of actual environments; thus, it is not rare that the action strategy obtained from the simulation makes no sense in practical applications. In this project, we developed a new framework, so-called data assimilation reinforcement learning (DARL) which incorporates data assimilation and reinforcement learning. DARL can provide the good action strategy in the small number of experiments by learning not only the action strategy but also the computer simulation simultaneously. We have also applied DARL to material design and drug discovery problems and confirmed its effectiveness compared with current methods.

研究分野：人工知能

キーワード：強化学習　データ同化　機械学習　人工知能

１．研究開始当初の背景

確率的な挙動を示すシステムにおいて，目的関数を最大化する最適な意思決定則(方策)を見つける最適意思決定問題は，最適制御，ファイナンス，データマイニング，オペレーションズ・リサーチなど幅広い領域に共通する研究課題である．この意思決定問題の解法の１つとして，近年，注目されているのが強化学習である．強化学習は現在の方策に従って実験をおこないデータを獲得し，得られたデータから方策を更新する．したがって，データさえ入手できればシステムが未知であっても適用できるため，これまで対処できなかった複雑な挙動を示す問題への応用が期待されている．

より複雑な意思決定問題への強化学習を展開する上で問題となるのはデータの収集である．複雑なシステムでデータを収集のための実験を行なうことは，経済的，時間的に無視できないコストが発生するため，学習に十分なデータ数を集めることができない．このデータ不足を回避するため，計算機シミュレーションを活用することは自然な発想である．しかしシミュレーションモデルに含まれるパラメータはおもにヒトの手で調整されるため,シミュレーションと実際のシステムの挙動が大きく乖離することは珍しくない.よって,シミュレーションで学習した方策が実際のシステム上で全く役に立たないという事態に陥ってしまう．したがって,強化学習においてシミュレーション を活用するためには,1) 数値モデルのパラメータを適切に決定し,シミュレーション精度を向上する方法が必要である.またシミュレーション精度の向上にも限界があるため,同時に2) シミュレーション誤差に対してロバストな強化学習法も併せて考える必要がある．

２．研究の目的

本研究の目的は，観測データから数値モデルのパラメータを学習するデータ同化を用いて，数値モデルを方策とともに学習するデータ同化強化学習を提案する．また，データ同化強化学習の応用としてシミュレーションを活用した創薬，新規材料の設計問題に取り組む．

３．研究の方法

(1)強化学習の数理基盤の再構築

数値シミュレーションのモデル化誤差に対してロバストな強化学習法を考えるためには，モデル化誤差が方策学習に与える統計的な影響を定量的に評価する必要がある．従来の強化学習の数理基盤は動的計画法(最適制御)を規範としているため，統計的な解析を行なうことは理論的な見通しがよくない．シミュレーションと強化学習の統合を 目指す本研究では,より簡潔でより拡張性の高い強化学習の枠組みを考えることが望まれる.申請者は先行研究(Ueno, T. et al., 2011, 2012)において統計学の観点から強化学習を考察し,方策学習を統計推論の問題として解こうと研究を続けてきた．この研究を自然な形で拡張し，方策学習を統計学における確率分布の推論問題として再定式化し，統計学で培われてきた洗練された解析法の適用を可能にする．また，解析結果から数値シミュレーションに対して頑健な方策学習法を提案する．

(2) シミュレーション精度の向上

データ同化は，システム同定の１手法であり，数値モデルのパラメータや初期条件をデータから学習することにより，数値シミュレーションの精度を向上させる枠組みである．近年，このデータ同化に高い表現能力を有することで知られるガウス過程を組み合わせた研究が提案され，高い評価を得ている．しかし，ガウス過程に基づく方法は，計算速度が遅いため，データ数の増加に対してスケールしない．本研究では，ランダム射影をもちいたカーネ

ル展開法を利用し，ガウス過程規範のデータ同化法の高速化を実現する．

(3)データ取得のスマート化

(2)で提案したガウス過程によるデータ同化法の強みの１つとして，シミュレーションによる予測結果だけでなく，その予測の信頼度(ばらつき)も知ることができる点である．この予測の信頼度を活用することにより，方策学習がより高速に学習できるようなデータ取得の最適化が可能となる．より具体的にウェブ広告，購買推薦システムなどで使われているベイズ最適化法を応用し，データ取得過程を制御し，少ない実験数で効率的に方策を学習できるようにする．

(4)新薬，新規材料開発への応用

データ同化強化学習の実応用として，新薬，新規材料設計問題に取り組む．新薬，新規材料設計の基礎段階として，複数の原子，分子，またはタンパク質を相互作用させて，所望の性質を満たす組み合わせを見つけなければならない．これは１回の実験に複数人で構成される研究者チームが数日，数週間かけて実験されることもあるなど，開発コストの負担の１つとなっている．これまで分子動力学シミュレーション，第一原理計算に基づく物理シミュレーションは研究されている．よってこれにデータ同化強化学習を適用することにより，候補から所望の性質を満たす組み合わせを効率良く見つけると期待できる．

4. 研究成果

(1)強化学習の数理基盤の再構築

方策を直接パラメトリックモデルであらわし，方策モデルのパラメータをデータから学習する新しい強化学習の枠組みを提案した．提案した枠組みでは方策パラメータの確率分布を導入し，その確率分布をカルバック・ライブラー距離（KL 距離）の最小化により学習する．これは方策学習を確率分布の推論問題に変換したことを意味し，これまで培われてきた洗練されたグラフィカルモデルの推論法を適用することを可能にする．提案法の理論的な性質，また実用上の強みを以下にまとめる．

① KL距離の最小化によって導かれた方策パラメータの確率分布は漸近的に最適なパラメータを中心としたデルタ分布に収束する．つまり，提案法は大域的に最適な方策に収束する．

② KL距離の最小化は解析的に計算できない場合が多いため，グラフィカルモデルの近似推論を用いる．この近似推論法を変更することにより，異なる性質をもつ強化学習法を導くことができる．より具体的にこれまで提案されてきた強化学習法のほとんどがこの近似推論を変更したものと解釈することができる．

③ 方策学習を確率推論問題に変換しているため，統計的な解析が行いやすい．特にシミュレーションを用いた場合の方策学習法は統計的にはパラメトリック推論，実データのみから学習する方法はセミパラメトリック推論と解釈できる．したがって，既存のパラメトリック推論の誤差に頑健なパラメータ学習法を用いることで，シミュレーション誤差に頑健な方策学習が実現できる．

この研究成果は研究成果，学会発表①，②にまとめられている．

(2)シミュレーション精度の向上

ガウス過程によるデータ同化は優秀であるが実験データの数に対して３乗の計算コストが必要となるため，スケーラビリティがない．これに対して，ランダム特徴射影を用いたガウス過程の近似法を提案した．この方法は，ガウス過程に必要な計算量をデータ数に対して線形まで減少させることが可能である．これにより，ガウス過程によるデータ同

化の著しい高速化に成功した．

(3)データ取得のスマート化

(2)で開発したガウス過程データ同化に基づくベイズ最適化法を提案した．より具体的にトンプソンサンプリング(Thompson，1933)を用いて実験クエリの生成法を最適化し，少ない実験数で効率良く方策学習を実現することに成功した．この手法の特筆すべき点として，従来の一般的なベイズ最適化法はデータ数に対して 2 乗の計算量を必要とするが，提案手法はデータに対して線形の計算量で計算ができるため，極めて高速に実験クエリを生成できる．したがって，大規模システムでこそ真価を発揮する方法であり，今後の拡張が見込まれる．この研究成果については，研究成果，雑誌論文①にまとめられている．

(4)新薬，新規材料開発への応用

研究成果(1)から(3)の結果を用いて，新規薬品，材料開発問題にデータ同化強化学習を応用した．具体的なタスクとしては，タンパク質・化合物ドッキングシミュレーションにおける最適な化合物の探索，モリブデンのナノクラスタ構造の最適化である．この両タスクとも正解が既知である問題にデータ同化強化学習を適用し，どれだけ効率よく見つけることができるか検証した．その結果，従来の探索法に比べて数十倍以上，場合によっては数百倍以上高速に所望の化合物，ナノクラスタをみつけることに成功した．今後，結果が未知の問題に適用し，その性能を検証する．

5．主な発表論文等

[雑誌論文](計5件)

① Ueno, T., Rhone, T.D., Hou, Z., Mizoguchi, T. and Tsuda, K., COMBO: An Efficient Bayesian Optimization Library for Materials Science, Materials Discovery, 2016, accepted．(査読有)

② 羽室 行信，植野 剛，鷲尾 隆．極大クリーク列挙技術のビジネス応用とソフトウェアツール,電子情報通信学会誌，vol．92，no．12，pp.1103-1106，2014．（査読無）URL: http://ci.nii.ac.jp/naid/11000989312

③ 植野 剛，前田 新一，川鍋 一晃．統計学習の観点から見た TD 学習 計測と制御，vol．52，no．3，pp．277-283，2013．（査読無）

④ 十河 泰弘，植野 剛，河原 吉伸，鷲尾 隆．Density power divergence を用いたロバスト能動回帰学習，人工知能学会論文誌，vol．28，no．1，pp．13-21，2013．（査読有）URL:http://doi.org/10.1527/tjsai.28.13

⑤ Sogawa Y., Ueno, T., Kawahara, Y., and Washio, T., Active learning for noisy oracle via density power divergence, Neural Networks (NNs), vol．46，pp．133-143，2013．(査読有) URL:http://europepmc.org/abstract/MED/23728156

[学会発表](計4件)

① 植野 剛．確率推論に基づく方策探索法，第 32 回 日本ロボット学会 学術講演会，2014年9月4日．(査読有)
会場名：九州産業大学(福岡県北九州市)

② 植野 剛．e 射影に基づく方策探索法,第 28 回人工知能学会全国大会，2014年5月13日．(優秀発表賞)（査読有）
会場名：ひめぎんホール(愛媛県松山市)

③ Ueno, T., Semiparametric Statistical Approach to Reinforcement Learning, *The ISI World Statistics Congress*, 2013年8月20日. (招待講演): 場所: 東京都文京区(東京大学)

④ 植野　剛．学習による制御，強化学習 鉄鋼学会：計測・制御・システム工学部会シンポジウム，2013年6月19日．(招待講演)
会場名：新日鉄千葉工場(千葉県千葉市)

６．研究組織

(1)研究代表者

植野 剛(UENO Tsuyoshi)
東京大学大学院新領域創成科学研究科
特任研究員
研究者番号：90615824