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　 ５ 月 １９ 日現在

機関番号：１２６０１
研究種目：基盤研究(A)（一般）
研究期間：2017～2021
課題番号：１７Ｈ００７５７

研究課題名（和文）統計的強化学習の深化と応用

研究課題名（英文）Theory and Application of Statistical Reinforcement Learning

研究代表者
杉山　将（Sugiyama, Masashi）

東京大学・大学院新領域創成科学研究科・教授

研究者番号：９０３３４５１５

交付決定額（研究期間全体）：（直接経費）　　34,600,000円

研究成果の概要（和文）：本研究では，逐次的意思決定および確率的推論の理論とアルゴリズム構築を行った．強化学習の研究では，実用性向上を目指し弱教師付き模倣学習や複雑な問題の階層化などの手法を開発し，その有効性を実験的に示した．多腕バンディット問題の研究では，線形バンディット，比較バンディット，良腕識別，組み合わせバンディットなどに対する理論保証付きアルゴリズムを開発した．確率的推論の研究では，ベイズ推論のロバスト化や近似計算の高速化，および，時間とともに発生する事象のモデル化に関する研究を行い，理論的・実験的に有効性を検証した．

研究成果の学術的意義や社会的意義
逐次的意思決定や確率的推論は，今後の発展が大いに期待される重要な機械学習技術である．本研究では，強化学習や多腕バンディットの適用範囲を拡大する新しいアルゴリズムを開発するとともに，確率的推論のロバスト性向上や近似計算の高速化に関する研究を行った．このような基礎理論的な研究成果は，逐次的意思決定や確率的推論の原理の解明に貢献するものであり，機械学習分野の主要国際会議で学術的に高い評価を受けた．また，開発したアルゴリズムの有効性は計算機実験によって示されており，将来の社会実装につながる社会的意義のある開発であるとも考えられる．

研究成果の概要（英文）：In this research, we developed theories and algorithms for sqeuential decision making and probabilistic inference. In the study of reinforcement learning, we developed methods for weakly supervised imitation learning and hierarchization of complex problems to improve their practicality, and demonstrated their effectiveness experimentally. For multi-arm bandit problems, we developed algorithms with theoretical guarantees for linear bandit, dueling bandit, good-arm identification, and combinatorial bandit. In the area of probabilistic inference, we have conducted research on making Bayesian inference robust, speeding up approximate computation, and modeling temporal events, and have verified the effectiveness of these methods both theoretically and experimentally.

研究分野：知能情報学

キーワード：強化学習　機械学習　多腕バンディット問題　模倣学習　ベイズ推論　ロバスト性

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

深層学習を用いた画像認識や言語翻訳が人間と同等以上の性能を達成し，囲碁などのゲームにおいては人工知能が人間のチャンピオンを打ち負かすレベルに到達した．しかし，将来的に機械学習技術をより複雑な実世界の問題で活用していくためには，更なる柔軟性と信頼性の向上が不可欠である．

２．研究の目的

本研究では，逐次的意思決定と確率的推論に関する基礎理論的な研究を行う．そして，学習の原理を理論的に解明するとともに実用的なアルゴリズムを開発し，その有効性を計算機実験によって示す．特に，従来技術では解決できない新しい問題の解法や，従来技術の信頼性を向上させる技術開発を行う．

３．研究の方法

逐次的意思決定に関しては，(a) マルコフ決定過程のもとでの強化学習問題，および，その一般化である模倣学習問題，更には，複雑な学習課題の階層化に対するアルゴリズム開発を行い，計算機実験によりその性能を評価する．また，(b) 状態遷移を持たない逐次的意思決定問題である多腕バンディット問題，および，その一般化問題に対して，理論保証付きアルゴリズムの開発を行う．(c) 確率的推論に関しては，ベイズ推論の枠組みのもとで異常値に対するロバスト化や近似計算の高速化，更には，時間とともに発生する事象のモデル化に関する研究を行い，その有効性を理論的および実験的に示す．

４．研究成果

(a) 強化学習

標準的な強化学習では報酬関数を事前に定義する必要があるが，現実問題では困難であることが多い．そこで，人間のエキスパートが効果的な方策を教師信号として提示し，それをもとに報酬関数を学習する模倣学習と呼ばれるアプローチが研究されている．しかし，エキスパートから十分な教師信号を得ることが難しいことが多いため，弱教師付き学習と呼ばれる強い教師信号を必要としない機械学習技術を用いた，独自の模倣学習のアルゴリズムを開発を行った．具体的には，教師データに信頼度が付与されている場合に不完全な教師信号からでも報酬関数を学習できる手法(Wu et al., ICML2019)，スキルにばらつきのあるエキスパートの集団から集めた教師データを効果的に活用できる変分推論型学習手法(Tangkaratt et al., ICML2020)，専門家と非専門家から得た教示データが混ざっている状況でのロバスト学習手法(Tangkaratt et al., AISTATS2021)，更には，ユーザが教師情報の取得プロセスに介入できる状況において苦手な部分の情報を積極的に取得する手法(Chen et al., ML2020)を開発した．

強化学習で解決したいタスクが非常に複雑な場合，それを複数の小タスクに分割することが有効である．階層型強化学習はそのような分割アプローチの一つであり，複数の方策を適用的に使い分ける階層型強化学習手法(Osa & Sugiyama, AAAI2018)，および，情報量最大化原理に基づく階層型強化学習のアルゴリズム(Osa et al., ICLR2019)を開発し，その有効性を計算機シミュレーションにより確認した．更に，強化学習の性能を様々な観点から向上させるべく，解の多様性を増すことによるロバスト化(Osa et al., NN2022)，連続時間の直接的扱いによる性能向上(Ohnishi et al., NeurIPS2018)，未知環境における安全性担保のための制約付き強化学習法(Ohnishi et al., Automatica2021)，価値関数の曲率情報を活用した高性能なアクター・クリティック法(Tangkaratt et al., ICLR2018)を開発した．

方策探索型の強化学習法では，勾配の期待値を標本から精度良く推定する事が重要である．一般的には，尤度比を用いた近似法と再パラメータ化を用いた近似法がこの目的に用いられるが，これらの手法の相互関係はよくわかっていない．そこで，これらの手法を含む統一的な枠組みを考案し，今後の勾配推定研究の基礎を築いた(Parmas & Sugiyama, AISTATS2021)．

(b) 多腕バンディット

報酬が線形関数で与えられる線形バンディット問題に対して，最適な選択肢を最小の回数で見つけられる理論保証付き最適腕識別アルゴリズムを開発した(Xu et al., AISTATS2018)．また，各選択肢から絶対的な報酬値が観測できない場合でも，二つの選択肢間の相対的な質的報酬比較だけからでも学習できる質的比較バンディット法のアルゴリズムを開発した(Xu et al., AAAI2019)．さらに，最適な選択肢を見つけるのに時間がかかる場合でも，「良い」

選択肢であれば素早く見つけられる良腕識別アルゴリズムを開発した(Kano et al, ML2019). また, トンプソン抽出とよばれる確率的なアルゴリズムを拡張し, 報酬が部分的にしか観測できない場合でも所望の性能が得られるアルゴリズムを構成した(Tsuchiya et al., NeurIPS2020).

組み合わせ的な選択肢から意思決定を行う組み合わせバンディット問題に対して, 個々の選択肢でなくそれらの集合からしか報酬が観測できない状況における最適腕探索問題を考え, ナイーブな解法では指数時間かかってしまうところを多項式時間で近似解が求められるアルゴリズムを開発した(Kuroki et al., NeCo2020). また, グラフにおける密サブグラフ発見問題に対して, 個々の枝でなく枝の集合からしか報酬を観測できない厳しい状況で, ほぼ最適な性能が得られる近似アルゴリズムを構成した(Kuroki et al., ICML2020).

(c) ベイズ推論

データに異常値が含まれる状況に対するベイズ推論問題に対して, 指数型分布族を拡張した q-指数型分布族を用いた期待値伝播アルゴリズムを開発した (Futami et al., NeurIPS2017). これは, モデルベースのロバスト推論手法と捉えることができる. 一方, モデルフリーのロバスト推論手法に関しても研究を行い, ロバスト汎距離を用いた変分学習手法を開発した(Futami et al., AISTATS2018). 更に, ロバスト汎距離を近似ベイズ計算に用いる計算技法も開発した(Fujisawa et al., AISTATS2021).

時間とともに発生する事象のモデル化に関して, ノンパラメトリックモデルを用いてイベントの発生時刻データをモデル化するベイズ推論アルゴリズム(Ding et al., AISTATS2018), および, 正確なイベントの発生時刻でなく時刻の区間しかわからない場合にも適用できる変分ベイズ推論アルゴリズム (Ding et al., UAI2018) を開発した.

更に, 複数の粒子を逐次的に更新していくことによって, 効率よく事後分布を近似する手法を開発した. 提案法では, フランク・ウルフ法を用いてカーネル空間における最大平均差異規準を最小化することによって, 効率よく事後分布を近似する(Futami et al., AAAI2019). 更に, 複数の粒子間の相互関係を考慮することによって, 更に効率よく事後分布を近似できる手法も開発した(Futami et al., ICML2020).

５．主な発表論文等

〔雑誌論文〕　計5件（うち査読付論文　5件／うち国際共著　2件／うちオープンアクセス　0件）

１．著者名 Kuroki Yuko、Xu Liyuan、Miyauchi Atsushi、Honda Junya、Sugiyama Masashi	４．巻 32
２．論文標題 Polynomial-Time Algorithms for Multiple-Arm Identification with Full-Bandit Feedback	５．発行年 2020年
３．雑誌名 Neural Computation	６．最初と最後の頁 1733～1773
掲載論文のDOI（デジタルオブジェクト識別子） 10.1162/neco_a_01299	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 Kano Hideaki、Honda Junya、Sakamaki Kentaro、Matsuura Kentaro、Nakamura Atsuyoshi、Sugiyama Masashi	４．巻 108
２．論文標題 Good arm identification via bandit feedback	５．発行年 2019年
３．雑誌名 Machine Learning	６．最初と最後の頁 721～745
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10994-019-05784-4	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 Chen Si-An、Tangkaratt Voot、Lin Hsuan-Tien、Sugiyama Masashi	４．巻 109
２．論文標題 Active deep Q-learning with demonstration	５．発行年 2019年
３．雑誌名 Machine Learning	６．最初と最後の頁 1699～1725
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10994-019-05849-4	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 該当する

１．著者名 Ohnishi Motoya、Notomista Gennaro、Sugiyama Masashi、Egerstedt Magnus	４．巻 127
２．論文標題 Constraint learning for control tasks with limited duration barrier functions	５．発行年 2021年
３．雑誌名 Automatica	６．最初と最後の頁 109504～109504
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.automatica.2021.109504	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 該当する

1．著者名 Osa Takayuki、Tangkaratt Voot、Sugiyama Masashi	4．巻 152
2．論文標題 Discovering diverse solutions in deep reinforcement learning by maximizing state?action-based mutual information	5．発行年 2022年
3．雑誌名 Neural Networks	6．最初と最後の頁 90～104
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.neunet.2022.04.009	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24件（うち招待講演　0件／うち国際学会　24件）

1．発表者名 Xu, L., Honda, J., & Sugiyama, M.
2．発表標題 Fully adaptive algorithm for pure exploration in linear bandits
3．学会等名 International Conference on Artificial Intelligence and Statistics (AISTATS2018)（国際学会）
4．発表年 2018年

1．発表者名 Ding, H., Khan, M. E., Sato, I., & Sugiyama, M.
2．発表標題 Bayesian nonparametric Poisson-process allocation for time-sequence modeling
3．学会等名 International Conference on Artificial Intelligence and Statistics (AISTATS2018)（国際学会）
4．発表年 2018年

1．発表者名 Tangkaratt, V., Abdolmaleki, A., & Sugiyama, M.
2．発表標題 Guide actor-critic for continuous control
3．学会等名 International Conference on Learning Representations (ICLR2018)（国際学会）
4．発表年 2018年

１．発表者名
Imamura, H., Sato, I., & Sugiyama, M.

２．発表標題
Analysis of minimax error rate for crowdsourcing and its application to worker clustering model

３．学会等名
International Conference on Machine Learning (ICML2018)（国際学会）

４．発表年
2018年

１．発表者名
Ding, H., Lee, Y., Sato, I., & Sugiyama, M.

２．発表標題
Variational inference for Gaussian process with panel count data

３．学会等名
Conference on Uncertainty in Artificial Intelligence (UAI2018)（国際学会）

４．発表年
2018年

１．発表者名
Ohnishi, M., Yukawa, M., Johansson, M., & Sugiyama, M.

２．発表標題
Continuous-time value function approximation in reproducing kernel Hilbert spaces

３．学会等名
Neural Information Processing Systems (NeurIPS2018)（国際学会）

４．発表年
2018年

１．発表者名
Tsuzuku, Y., Sato, I., & Sugiyama, M.

２．発表標題
Lipschitz-margin training: Scalable certification of perturbation invariance for deep neural networks

３．学会等名
Neural Information Processing Systems (NeurIPS2018)（国際学会）

４．発表年
2018年

１．発表者名
Futami, F., Cui, Z., Sato, I., & Sugiyama, M.

２．発表標題
Bayesian posterior approximation via greedy particle optimization

３．学会等名
AAAI Conference on Artificial Intelligence (AAAI2019)（国際学会）

４．発表年
2019年

１．発表者名
Xu, L., Honda, J., & Sugiyama, M.

２．発表標題
Dueling bandits with qualitative feedback

３．学会等名
AAAI Conference on Artificial Intelligence (AAAI2019)（国際学会）

４．発表年
2019年

１．発表者名
Xu, L., Honda, J., & Sugiyama, M.

２．発表標題
Fully adaptive algorithm for pure exploration in linear bandits

３．学会等名
International Conference on Artificial Intelligence and Statistics (AISTATS2018)（国際学会）

４．発表年
2018年

１．発表者名
Futami, F., Sato, I., & Sugiyama, M.

２．発表標題
Expectation propagation for t-exponential family using q-algebra

３．学会等名
Neural Information Processing Systems (NeurIPS2017)（国際学会）

４．発表年
2017年

１．発表者名
Futami, F., Sato, I., & Sugiyama, M.

２．発表標題
Variational inference based on robust divergences

３．学会等名
International Conference on Artificial Intelligence and Statistics (AISTATS2018)（国際学会）

４．発表年
2018年

１．発表者名
Ding, H., Khan, M. E., Sato, I., & Sugiyama, M.

２．発表標題
Bayesian nonparametric Poisson-process allocation for time-sequence modeling

３．学会等名
International Conference on Artificial Intelligence and Statistics (AISTATS2018)（国際学会）

４．発表年
2018年

１．発表者名
Osa, T. & Sugiyama, M.

２．発表標題
Hierarchical policy search via return-weighted density estimation

３．学会等名
AAAI Conference on Artificial Intelligence (AAAI2018)（国際学会）

４．発表年
2018年

１．発表者名
Tangkaratt, V., Abdolmaleki, A., & Sugiyama, M.

２．発表標題
Guide actor-critic for continuous control

３．学会等名
International Conference on Learning Representations (ICLR2018)（国際学会）

４．発表年
2018年

１．発表者名
Tangkaratt, V., Han, B., Khan, M. E., & Sugiyama, M.

２．発表標題
Variational imitation learning with diverse-quality demonstrations

３．学会等名
Proceedings of 37th International Conference on Machine Learning (ICML2020)（国際学会）

４．発表年
2020年

１．発表者名
Kuroki, Y., Miyauchi, A., Honda, J., & Sugiyama, M.

２．発表標題
Online dense subgraph discovery via blurred-graph feedback

３．学会等名
Proceedings of 37th International Conference on Machine Learning (ICML2020)（国際学会）

４．発表年
2020年

１．発表者名
Futami, F., Sato, I., & Sugiyama, M.

２．発表標題
Accelerating the diffusion-based ensemble sampling by non-reversible dynamics

３．学会等名
Proceedings of 37th International Conference on Machine Learning (ICML2020)（国際学会）

４．発表年
2020年

１．発表者名
Tsuchiya, T., Honda, J., & Sugiyama, M.

２．発表標題
Analysis and design of Thompson sampling for stochastic partial monitoring

３．学会等名
Neural Information Processing Systems (NeurIPS2020)（国際学会）

４．発表年
2020年

１．発表者名
Osa, T., Tangkaratt, V., & Sugiyama, M.

２．発表標題
Hierarchical reinforcement learning via advantage-weighted information maximization

３．学会等名
Proceedings of Seventh International Conference on Learning Representations (ICLR2019)（国際学会）

４．発表年
2019年

１．発表者名
Wu, Y.-H., Charoenphakdee, N., Bao, H., Tangkaratt, V., & Sugiyama, M.

２．発表標題
Imitation learning from imperfect demonstration

３．学会等名
Proceedings of 36th International Conference on Machine Learning (ICML2019)（国際学会）

４．発表年
2019年

１．発表者名
Parmas, P. & Sugiyama, M.

２．発表標題
A unified view of likelihood ratio and reparameterization gradients

３．学会等名
International Conference on Artificial Intelligence and Statistics (AISTATS2021)（国際学会）

４．発表年
2021年

１．発表者名
Tangkaratt, V., Charoenphakdee, N., & Sugiyama, M.

２．発表標題
Robust imitation learning from noisy demonstrations

３．学会等名
International Conference on Artificial Intelligence and Statistics (AISTATS2021)（国際学会）

４．発表年
2021年

１．発表者名
Fujisawa, M., Teshima, T., Sato, I., & Sugiyama, M.

２．発表標題
γ-ABC: Outlier-robust approximate Bayesian computation based on a robust divergence estimator

３．学会等名
International Conference on Artificial Intelligence and Statistics (AISTATS2021)（国際学会）

４．発表年
2021年

〔図書〕　計1件

〔産業財産権〕

〔その他〕

論文リスト
http://www.ms.k.u-tokyo.ac.jp/sugi/publications.html

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
研究協力者	長　隆之 (Osa Takayuki)		
研究協力者	タンカラット　ブット (Tangkaratt Voot)		
研究協力者	本多　淳也 (Honda Junya)		

6．研究組織（つづき）

	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
研究協力者	佐藤　一誠 (Sato Issei)		

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