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

平成　３０　年　　６　月　　７　日現在

機関番号：　２４４０３
研究種目：　基盤研究(C)（一般）
研究期間：　2015～2017
課題番号：　１５Ｋ００３４４

研究課題名（和文）漸近最適戦略の動的適応学習アルゴリズムへの応用

研究課題名（英文）Application of Asymptotic Optimal Strategy to Dynamic Adaptive Learning Algorithm

研究代表者

野津　亮（Notsu, Akira）

大阪府立大学・人間社会システム科学研究科・准教授

研究者番号：４０４０５３４５

交付決定額（研究期間全体）：（直接経費）　　3,500,000円

研究成果の概要（和文）：本課題では強化学習や最適化問題において確率論的に最適な選択を行うための方法について研究を進めた．選択肢が複数ある場合，過去の経験がどれだけあるか，良い結果がどれくらい見込めるかを基づいて判断する必要がある．本研究では強化学習や最適化問題においても同様であることを確認しつつ，最適な戦略を導入するための枠組みをいくつか考案することができた．特に，ベイズ推定の観点から強化学習アルゴリズム根本的に見直し，再構築できたことは学習と意思決定を切り分ける従来の一般的な考え方に一石を投じるものであると考えている．また，学習主体の状態推定を計算負荷をかけずに行う方法についても研究成果をあげることができた．

研究成果の概要（英文）：In this subject, we have studied a method for stochastically optimal selection in reinforcement learning and optimization problems. When there are multiple choices, it is necessary to judge based on how much past experience and how much good results can be expected. In this research, we were able to devise several frameworks for introducing the optimal strategy while confirming that it is the same in reinforcement learning and optimization problems. In particular, from the viewpoint of Bayesian estimation, the reinforcement learning algorithm was fundamentally reviewed and the reconstruction showed that the conventional general idea of separating learning from decision making was wrong. In addition, we also gave research results on the method of estimating the state of the learner without applying computational load.

研究分野：　ソフトコンピューティング

キーワード：　強化学習　最適化問題　漸近最適戦略　自己組織化マップ　意思決定　クラスタリング

１．研究開始当初の背景

漸近最適戦略に関する基礎的な研究としては，UCB1-tuned のように価値推定値の分散も利用しながら最大損失を推定する手法や，動的な環境に対する学習方法として，重み付き平均を価値推定値とするもの，一部の環境がスワップするような問題に特化したメタな学習機構を備えたものなど，さまざまなものが提案されてきていた．応用研究としては，ゲーム AI に利用する手法が有名である．UCT 探索によってコンピュータ囲碁などのアルゴリズムは急速に発展してきた．また，強化学習における状態数の爆発に対して，関数近似や様々なクラスタリングアプローチが近年盛んに研究されている．価値推定値の低い状態を削除するものや似た状態を統合したりするものなど，問題に応じた様々なものがある．いずれも国内外で活発に研究されている．

２．研究の目的

本研究では多腕バンディット問題最適化アルゴリズムである UCB (Upper Confidence Bound) 手法などに代表される漸近最適戦略とデータマイニングの分野でデータ構造を抽出するために用いられる共クラスタリング技術を知的エージェントの学習に応用する．漸近最適戦略は，価値発掘と価値追及をバランスよく選択するための手法だが，Q 学習などの強化学習, DE や PSO などの最適化手法においても，行動選択における探索（価値追求）と開拓（価値発掘）のバランスをどのようにとるかは非常に大きな問題である．本研究は主に強化学習と最適化問題における探索と開拓のバランス最適化アルゴリズムを開発し，共クラスタリング技術によってより適応的な環境理解を可能とした学習システムにそれを組み込み，新たな知的情報処理エージェントを提案する．

３．研究の方法

本研究では，漸近最適戦略や共クラスタリング技術を知的エージェントで効果的に利用するための条件と，二つの技術を融合することで生まれる新しい学習アルゴリズムの有用性を明らかにする．漸近最適戦略，共クラスタリング技術をそのまま知的エージェント技術に適用することの効果については，すでにいくつか研究成果発表しており，強化学習エージェントに漸近最適戦略を行動決定手法として適用することによって効率的な学習ができること，あるいは学習中の価値推定値テーブルを用いて，状態と行動を共クラスタリングすることで学習速度を早められることなどを確認してきた．本課題では対象とする問題に即した戦略や改良を加えたものの能力について，統計学的基盤を与え，各種シミュレーション実験を行う．また，新たに最適化問題へ漸近最適戦略とクラスタリング技術を適用する．最適化問題では探索（価値追求）と開拓（価値発掘）のバランスはパラメータやアルゴリズムに組み込まれていることが大半であり，確率的な収束性は論じられているものの，その期待値について言及しているものは少ない．本課題ではグリッド空間上での解探索がグリッドの選択という点で多腕バンディット問題となることを示し，UCB 手法等を適用することでリグレット（期待損失）最適化という側面で優位性を持つ方法になることを明らかにする．

漸近最適戦略と共クラスタリング技術を融合させる研究では，設定次第でクラスタリングによる情報量の圧縮とリグレット最適化が安定的に両立可能であることを明らかにする．共クラスタリング技術を強化学習に適用する際には，行動決定手法が大きな影響力を持つことがわかっており，状態行動空間におけるクラスター抽出を漸近最適戦略で安定化させることができると考えている．過去に発表した論文では，学習中のクラスター抽出のタイミングを価値推定値のエントロピーを計算することで決定することができると明らかにしたが，新しくリグレット値の評価も加えることで，より適切な分割タイミングを探る．

４．研究成果

UCT アルゴリズムの他に，最も効果的な漸近最適戦略の一つであるトンプソンサンプリングを強化学習アルゴリズムに組み込み，特に複雑な報酬環境で効果的な強化学習手法であることを確認した．正の報酬と負の報酬の学習という二点のみを用いた学習アルゴリズムで，認知モデルとしてのシンプルさを保ちながら，高度な意思決定が求められる環境でも学習が可能である．これについて論文発表を行った．

また，共クラスタリング技術を強化学習に応用するにあたり，いくつかの手法を検討したが，自己組織化マップを用いる方法が可視化に適しており，かつ，学習の誤差も少なくなることを確認した．学習環境が複雑になると一般的に計算量が爆発的に増えていくが，強化学習における状態空間のクラスタリングについて，成長型自己組織化マップをもちいることによってオンライン型で速く学習させることができることを発見できた．従来のクラスタリング技術では学習中にクラスタリングを適用すると，新しい状態が観測されたときにそれまでの学習結果が壊れてしまうことがあったが，この成長型のアルゴリズムによって既存の学習結果を壊さずに状態空間を定義し，速い学習速度を保持することができ，かつ，必要最低限の計算量やメモリの確保で学習できることが確認できた．基礎的な研究については国内発表を終え，論文投稿を行った．

また，漸近最適戦略を差分進化アルゴリズムに応用する研究を進めた．これは最適化問題における探索アルゴリズムの探索と活用のバランスを改善し，探索効率を大幅に改善するものである．従来法は初期探索の効率が悪いが，これは，次の探索点を決める際に良い

解が得られそうなところを探索するのか，新しい情報を求めて別な場所を探索するのかの調節方法について統計的な視点からアルゴリズム化されていないためである．差分進化アルゴリズムをUCTアルゴリズムと融合させ，両者の長所を兼ね備えた探索アルゴリズムを提案した．
研究期間全体を通じ，状態・行動空間のクラスタリングと漸近最適戦略により，オンライン型強化学習アルゴリズムを大きく発展させることができた．近年，バッチ型の強化学習が注目を浴びることが多いが，オンライン型で柔軟で，計算量が少なく高価な計算機を必要としない学習アルゴリズムを開発することによって，機械学習の適用範囲を大きく広げることができた．また，人間の持つしなやかで適応的な学習能力を再現する一つの認知モデルを提案できたことは心理学的にも重要な意味を持っていると考えている．

５．主な発表論文等
（研究代表者、研究分担者及び連携研究者には下線）

〔雑誌論文〕（計１１件）


1. Deterministic Annealing Process for pLSA-induced Fuzzy Co-clustering and Cluster Splitting Characteristics
T. Goshima, K. Honda, S. Ubukata, A. Notsu
International Journal of Approximate Reasoning, 95, 185-193 (2018).
doi: 10.1016/j.ijar.2018.02.005
査読あり

2. Designation of Candidate Solutions in Differential Evolution Based on Bandit Algorithm
M. Sakakibara, A. Notsu, S. Ubukata, K. Honda
Proc. of the 18th International Symposium on Advanced Intelligent Systems, #F1c-2, 471-478 (2017).
査読あり

3. Phase Transition in pLSA-induced Fuzzy Co-clustering Based on Tuning of Intrinsic Fuzziness
T. Goshima, K. Honda, S. Ubukata, A. Notsu
Proc. of the 18th International Symposium on Advanced Intelligent Systems, #T2c-1, 243-249 (2017).
査読あり

4. Visual Assessment of Co-cluster Structure through Cooccurrence-Sensitive Ordering
K. Honda, T. Sako, S. Ubukata, A. Notsu
Proc. of Joint 17th World Congress of International Fuzzy Systems Association and 9th International Conference on Soft Computing and Intelligent Systems, #50, 1-6 (2017).
査読あり

5. FCM-type Fuzzy Coclustering for Three-mode Cooccurrence Data: 3FCCM and 3Fuzzy CoDoK
K. Honda, Y. Suzuki, S. Ubukata, A. Notsu
Advances in Fuzzy Systems, 2017, #9842127, 1-8 (2017).
DOI:10.1155/2017/9842127
査読あり

6. プロスペクト理論を応用したベータ分布伝搬型強化学習による効率的探索と活用
野津 亮，生方誠希，本多克宏
知能と情報(日本知能情報ファジィ学会誌)，29，1，507-516 (2017).
DOI: 10.3156/jsoft.29.1_507
査読あり

7. Visualization of Learning Process in “State and Action” Space Using Self-Organizing Maps
A. Notsu, Y. Hattori, S. Ubukata, K. Honda
Journal of Advanced Computational Intelligence and Intelligent Informatics, 20, 6, 983-991 (2016).
doi: 10.20965/jaciii.2016.p0983
査読あり

8. Application of the UCT Algorithm for Noisy Optimization Problems
A. Notsu, S. Kane, S. Ubukata, K. Honda
Proc. of Joint 8th International Conference on Soft Computing and Intelligent Systems and 17th International Symposium on Advanced Intelligent Systems, 48-53 (2016).
査読あり

9. バンディットアルゴリズムに基づいた汎用最適化手法の開発
野津 亮，河上 寛和，本多克宏，生方誠希
知能と情報（日本知能情報ファジィ学会誌），28，1，522-534 (2016).
doi: 10.3156/jsoft.28.522
査読あり

10. Performance Investigation of UCB

Policy in Q-Learning
K. Saito, A. Notsu, S. Ubukata, K. Honda
Proc. of 14th International Conference on Machine Learning and Applications, 770-773 (2015).
査読あり

11. Proposal of Grid Area Search with UCB for Discrete Optimization Problem
A. Notsu, K. Saito, Y. Nohara, S. Ubukata, K. Honda
Proc. of Integrated Uncertainty in Knowledge Modelling and Decision Making, LNCS 9376, 102-111 (2015).
査読あり

〔学会発表〕（計　8件）

1. 第 61 回システム制御情報学会研究発表講演会
榊原 雅也, 野津 亮, 生方 誠希, 本多 克宏
バンディットアルゴリズムに基づいた差分進化における解集団の生成
講演論文集, #343-1, 1-4, 2017.

2. 第 33 回ファジィシステムシンポジウム
野津 亮, 柳川 綾香, 生方 誠希, 本多 克宏
トンプソンサンプリングにおけるサンプリングの省略
講演論文集, #FF1-1, 661-664, 2017.

3. 第 32 回ファジィシステムシンポジウム
菊田 美月, 野津 亮, 生方 誠希, 本多 克宏
認知特性に基づいたバンディットアルゴリズムの頑強性
講演論文集, WF3-4, 283-288, 2016.

4. 第 26 回インテリジェント・システム・シンポジウム
野津 亮, 近藤 佑紀, 生方 誠希, 本多 克宏
自己組織化マップを用いた強化学習結果の抽象化とその利用
講演論文集, F2A1, 126-129, 2016.

5. 第 31 回ファジィシステムシンポジウム
服部 雄市, 野津 亮, 生方 誠希, 本多 克宏, 上野 貴紀
Q 学習におけるファジィ共クラスタリングによる知識の圧縮と再利用
講演論文集, WD3-1, 199-204, 2015.

6. 第 31 回ファジィシステムシンポジウム
斉藤 晃貴, 野津 亮, 野原 由布美, 生方 誠希, 本多 克宏
UCB による離散最適化問題の探索と活用の調整
講演論文集, WE2-2, 240-245, 2015.

7. 第 25 回インテリジェント・システム・シンポジウム
斉藤 晃貴, 野津 亮, 生方 誠希, 本多 克宏
Q 学習における UCB 行動選択手法の性能に関する調査
講演論文集, 148-153, 2015.

8. 第 25 回インテリジェント・システム・シンポジウム
服部 雄市, 野津 亮, 生方 誠希, 本多 克宏
強化学習における自己組織化マップを用いた状態と行動の学習プロセスの可視化
講演論文集, 148-153, 2015.

〔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：
出願年月日：
国内外の別：

○取得状況（計　　件）

名称：
発明者：
権利者：
種類：
番号：
取得年月日：
国内外の別：

〔その他〕
ホームページ等
http://www.cs.osakafu-u.ac.jp/hi/

## ６．研究組織

(1)研究代表者
野津　亮　　（Notsu Akira）
大阪府立大学・人間社会システム科学研究科・准教授
研究者番号：40405345

(2)研究分担者
本多　克宏　（Honda Katsuhiro）
大阪府立大学・工学研究科・教授
研究者番号： 80332964

(3)連携研究者
（　　）

研究者番号：

(4)研究協力者
（　　）