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　５ 年　 ６ 月　 １ 日現在

機関番号：２４４０５
研究種目：基盤研究(C)（一般）
研究期間：2018～2022
課題番号：１８Ｋ１１４７３

研究課題名（和文）完全オンライン型強化学習システムにおける時間と空間の分節化

研究課題名（英文）Segmentation of Time and Space in a Fully Online Reinforcement Learning System

研究代表者
野津　亮（Notsu, Akira）

大阪公立大学・大学院現代システム科学研究科　・教授

研究者番号：４０４０５３４５

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）：成長型自己組織化マップを強化学習向けに改良し，学習効率を維持しながらも状態空間・状態遷移を教師無し学習で学習する手法を考案し，その有用性を湿すことができた．また，ハイパーパラメータの設定を適応的に大きく変更することでも，環境に適応できることを示した．さらに，最適化アルゴリズム手法の中でも優秀な差分進化に対し，いくつかの局所的環境を推定しながら手法を切り替える方法を提案し，性能を改善することができた．加えて，当初はあまり検討していなかった深層強化学習へ本研究で得られた知見を応用し，全く新しい深層強化学習システムを提案できた．

研究成果の学術的意義や社会的意義
本研究は強化学習が必要とする空間を統計学的に大量のデータを用いて獲得するのでは無く，幾何学的なミクロな観点から獲得したという意味で学術的な意義があると考えている．また，機械学習にとってハイパーパラメータの設定は大きな問題であるが，その適応的変化や並列学習で対応できることを明らかにしたことは，学術的にも産業応用を考えた上でも意義がある．さらに，ブラックボックス最適化アルゴリズムを発展させることは複雑化する社会問題など，ありとあらゆる最適化に貢献できることを意味しているので，社会的にも大きな意義がある．

研究成果の概要（英文）：We modified the growing self-organizing map for reinforcement learning and devised a method for unsupervised learning of state space and state transitions while maintaining learning efficiency, and demonstrated the usefulness of this method. We also showed that the method can adapt to the environment by adaptively changing the hyperparameter settings significantly. Furthermore, we proposed a method for switching methods while estimating several local environments for differential evolution, which is one of the best optimization algorithm methods, and were able to improve the performance. In addition, we were able to apply our findings to deep reinforcement learning, which had not been considered much at first, and propose a completely new deep reinforcement learning system.

研究分野：強化学習

キーワード：強化学習　クラスタリング　最適化アルゴリズム　転移学習　学習と進化

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

強化学習研究において,「状態空間，行動空間をどのように定義するか」と「計算効率をどのように上げるか」が大きな研究課題であった．状態遷移における各状態の定義とその遷移関係を適切に推定するには多くの情報と最適化計算が必要で,実機への応用は難しいと考えていた.また,スパースな報酬空間に対しては内発的な報酬を定義することでクリア可能であることは我々の研究を含めいくつか示されてきたが，それも状態の定義をどのようにするかに大きく依存している．

また,深層強化学習による関数近似,内発的報酬も状態定義に関する情報を与えておく必要があった．基本的には,内発的な報酬を得るために今の状態に過去何回訪れてきたかをカウントすることが必要なことであるのだが，状態が同じか違うかに関するヒントを学習エージェントに与えることは，環境に対する情報を事前に得ておく必要があった．

さらに,時間とか学習周期もどのように定義するかもこれらの問題と関係している．学習・行動変化のタイミングを細かく設定すれば考慮すべき状態数が増加し，学習が困難になることは明らかである．逆に,周期が大きくなると細かい制御や学習ができなくなって問題に対応できなかったり，低い報酬しか得られないといった問題が生じる．

２．研究の目的

本研究の目的は大きく三つあった．一つは学習効率を維持しながらも状態空間・状態遷移を教師無し学習で取得することであり，もう一つは学習周期などのハイパーパラメータを同時に学習させる方法の確立，さらには，学習において状態遷移による非連続な価値の変化に対応できる,局所解に陥りにくい最適化アルゴリズムを開発することであった．

状態空間・状態遷移をシンプルな教師無し学習で分節化し獲得することで,学習に適した状態空間を計算コストを抑えつつ獲得する．これによって,実機への応用可能性を高め,学習エージェントをスマートフォンなどの身近なものに組み込むことができると考えている．また,データの密度による状態空間の分節化は状態遷移の獲得という点では不適切だったので，これまでの一般的な分節化アルゴリズムとは違ったかたちの新しいクラスタリング技術の開発する．

学習周期などの時間を分節化する手法を検討することで,これが自動化できるのか,そうで無いのかが明らかになる．加えて，これらハイパーパラメータを変更するにあたり,学習を最初からやり直す必要があるのかどうか,やり直さない場合,どのような影響があるかを明確にすることで，計算コストを考慮に入れた最適化も検討可能になると考えられた．

局所解に陥りにくい最適化アルゴリズムを開発することは最適化アルゴリズム開発における基本的な課題であるので，これは直接的にブラックボックス最適化アルゴリズム研究に貢献するものである．一方で,近年の最適化アルゴリズム研究が対象としているような超高次元の最適化問題を強化学習におけるハイパーパラメータ設定問題とするのには隔たりがあるので，比較的低次元で，解空間の局所的・幾何学的な特徴を利用するような手法の開発が目標となる．

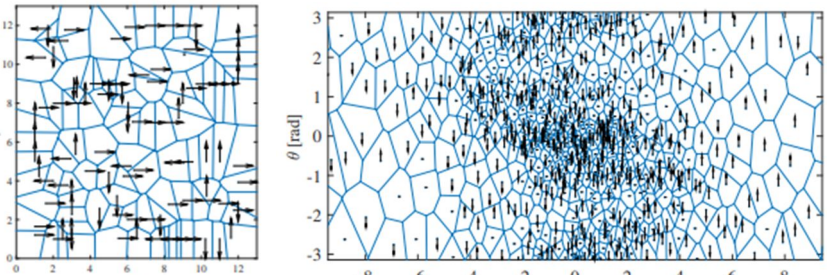
３．研究の方法

研究方法は主に文献調査，手法の検討，シミュレーション実験からなる．文献調査では主要な論文誌，国際会議の強化学習，最適化アルゴリズムに関する論文を参照し，本研究課題に応用・流用できるかを判断するための資料を収集する．次に，それら資料から検討できることと議論の中で生まれた本質的な問い（例えば，密度に依存しないクラスタリング技術の開発方法は可能かなど）から，新たな手法を考案する．最後にコンピュータシミュレーション実験によってその効果を確認した．

４．研究成果

研究成果として，上記の三つの目標の達成と新たに深層強化学習へ手法を適用し，成果を出すことができた点を挙げることができる．まず，成長型自己組織化マップを強化学習向けに改良し,学習効率を維持しながらも状態空間・状態遷移を教師無し学習で学習する手法を考案し，その有用性を示すことができた.また,ハイパーパラメータの設定を適応的に大きく変更することでも,環境に適応できることを示した．さらに,最適化アルゴリズム手法の中でも優秀な差分進化に対し,いくつかの局所的環境を推定しながら手法を切り替える方法を提案し,性能を改善することができた．加えて,当初はあまり検討していなかった深層強化学習へ本研究で得られた知見を応用し，全く新しい深層強化学習システムを提案できたことは大きな成果だと考えている．

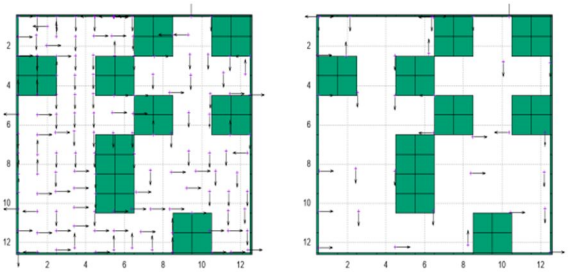
成長型自己組織化マップによる強化学習システムでは，一般的な深層強化学習手法では不可能な問題に対しても自律的に空間を分割して学習を進めることができる手法を提案できた．通常は上でも記述したように，密度によって状態・クラスタを定義するが，これをそのまま強化学習に適用すると，クラスタ中心が動きすぎて状態遷移関係を報酬値の大小関係によって上手く捉えることが困難になってしまうことがわかった．そこで,クラスタ中心がなるべく変化しないよ



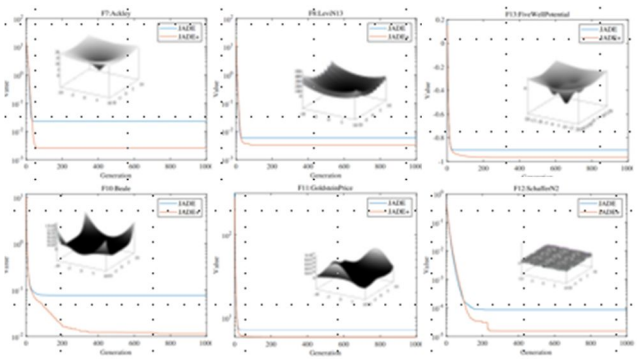
成長型自己組織化マップによる状態空間の分節化の例

う，かつ新たな状態を動的に獲得するためのアルゴリズムのコアなアイディアとして，入力値の差分に着目した．新たな状態・クラスタを作るかどうかをパラメータとして定義してしまうのでは無くて，入力値の差分を学習する機構を導入して，それとの比率をクラスタの作成基準とした．これによって効率的な状態空間を構築しつつ学習が可能になった．

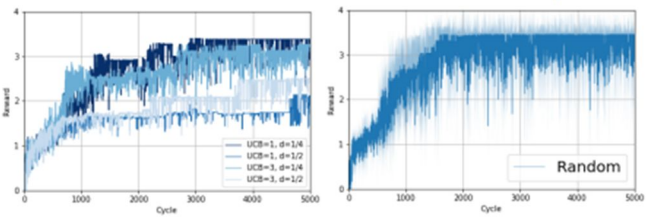
ハイパーパラメータの一つである学習周期を適応的に変化させることの影響について，想像以上に柔軟に対応できることを明らかにした．例えば，学習が上手くいかなかったときに，学習周期を減少，増加，ランダムに変更するというような手法を検討したが，黄金分割法，ベイズ最適化のような高度な手法を用いる手法と比べてそれほど性能が低下しないことが分かった．これは，オンライン学習であれ，過去の情報が価値推定値として保存され，パラメータが変わったとしても経験として機能するということであった．もちろん，状態遷移が変化するのでどの程度過去の経験が生きるか，価値推定値の更新が難しくなるかは問題依存ではあるが，少なくとも迷路の学習や倒立振り子の問題などのような基本的なベンチマークテストのような学習では問題無く動作した．さらにこの知見は深層強化学習の並列変へ応用することができた．



ハイパーパラメータの適応的変化と学習結果の例



JADE にネルダーミード法の要素を追加し性能向上



複数の深層強化学習（左）と提案法で統合された深層強化学習（右）の報酬値

また，解空間の幾何学的な特徴を考慮した最適化アルゴリズムの改良にも取り組み成果を上げた．具体的には，局所解探索に強いネルダーミード法の良い点を上手く差分進化手法に組み込み直すことで性能を向上させた．本研究で提案した方法はいくつかあるが，基本的な考え方としては，局所解に陥っているかどうかを解の更新比率・解の空間分布と最良解の位置などから推定し，集団の外側を必要な時には外側への探索を追加したり，内側に解があると推定できるときにはネルダーミード法に切り替えるというものであった．

さらに，当初はあまり検討していなかった，深層強化学習へベイズ推定的な内発的報酬の追加とそれのパラメータの適応的最適化研究も行った．経験度によるペナルティをサイクル数から定義することで，学習範囲を経験があるかないかのみに絞り，強化学習における探索と活用のバランス化を図った．さらに，それらのパラメータを含め何パターンかの深層強化学習ネットワークを並列的に学習させ，そのなかの良いものを選択させることを実験した．結果としては各ネットワークの意思決定が相互に良い学習経験を提供し，学習範囲を限定・誘導し，効率的な学習をするネットワーク群・集団として機能したことが明らかになった．

以上のように本研究では，強化学習に関連する重要な状態空間の定義，ハイパーパラメータの設定手法，最適化について，性能面からも学術的にも興味深い知見を得ることができたと考えている．これらの研究は次の研究課題へと引き継いで，研究を続けていく．

５．主な発表論文等

〔雑誌論文〕　計10件（うち査読付論文　10件／うち国際共著　0件／うちオープンアクセス　2件）

１．著者名 K. Yasunaga, A. Notsu, S. Ubukata, K. Honda	４．巻 G01-2
２．論文標題 A Study on Pre-Learning of State Similarity for Deep Reinforcement Learning	５．発行年 2021年
３．雑誌名 Proc. of 22nd International Symposium on Advanced Intelligent Systems	６．最初と最後の頁 7-16
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 Y. Miyahira, A. Notsu	４．巻 9
２．論文標題 Additional Out-group Search for JADE	５．発行年 2022年
３．雑誌名 Integrated Uncertainty in Knowledge Modelling and Decision Making	６．最初と最後の頁 105-116
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-98018-4	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 A. Notsu, K. Yasuda, S. Ubukata, K. Honda	４．巻 97
２．論文標題 Online state space generation by a growing self-organizing map and differential learning for reinforcement learning	５．発行年 2020年
３．雑誌名 Applied Soft Computing	６．最初と最後の頁 1-9
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.asoc.2020.106723	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 －

１．著者名 J. Tsubamoto, A. Notsu, S. Ubukata, K. Honda	４．巻 1
２．論文標題 Proposal of Adaptive Randomness in Differential Evolution	５．発行年 2020年
３．雑誌名 Proc. of 2020 IEEE Congress on Evolutionary Computation	６．最初と最後の頁 1-8
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 A. Notsu, J. Tsubamoto, Y. Miyahira, S. Ubukata, K. Honda	4．巻 1
2．論文標題 Randomness Selection in Differential Evolution Using Thompson Sampling	5．発行年 2020年
3．雑誌名 Proc. of Joint 11th International Conference on Soft Computing and Intelligent Systems and 21st International Symposium on Advanced Intelligent Systems	6．最初と最後の頁 351-355
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 M. Sakakibara, A. Notsu, S. Ubukata, K. Honda	4．巻 23
2．論文標題 Designation of Candidate Solutions in Differential Evolution Based on Bandit Algorithm and its Evaluation	5．発行年 2019年
3．雑誌名 Journal of Advanced Computational Intelligence and Intelligent Informatics	6．最初と最後の頁 758-766
掲載論文のDOI（デジタルオブジェクト識別子） 10.20965/jaciii.2019.p0758	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 －

1．著者名 A. Notsu, K. Yasuda, K. Yasunaga, S. Ubukata, K. Honda	4．巻 1
2．論文標題 Simple Pre-learning of State Similarity for Deep Reinforcement Learning	5．発行年 2019年
3．雑誌名 Proc. of the 20th International Symposium on Advanced Intelligent Systems and 2019 International Conference on Biometrics and Kansei Engineering	6．最初と最後の頁 63-68
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 J. Tsubamoto, A. Notsu, S. Ubukata, K. Honda	4．巻 1
2．論文標題 Changes during the Search from Differential Evolution with Intervals to Other Methods	5．発行年 2019年
3．雑誌名 Proc. of the 20th International Symposium on Advanced Intelligent Systems and 2019 International Conference on Biometrics and Kansei Engineering	6．最初と最後の頁 75-80
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 A. Notsu, K. Yasuda, S. Ubukata, K. Honda	4．巻 #12428
2．論文標題 Optimization of Learning Cycles in Online Reinforcement Learning Systems	5．発行年 2018年
3．雑誌名 Proc. of 2018 IEEE International Conference on Systems, Man, and Cybernetics	6．最初と最後の頁 3520-3524
掲載論文のDOI（デジタルオブジェクト識別子） なし	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 A. Notsu, M. Sakakibara, S. Ubukata, K. Honda	4．巻 #T1a-2
2．論文標題 Setting of Candidate Solutions Considering Confidence Intervals in Differential Evolution	5．発行年 2018年
3．雑誌名 Proc. of 2018 International Conference on Fuzzy Theory and Its Applications	6．最初と最後の頁 7-11
掲載論文のDOI（デジタルオブジェクト識別子） なし	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12件（うち招待講演　1件／うち国際学会　0件）

1．発表者名 宮平 裕一，野津 亮，本多 克宏，生方 誠希
2．発表標題 差分進化におけるパラメータのバンディットアルゴリズムによる適応的選択
3．学会等名 第65回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 安永 恭平，野津 亮，生方 誠希，本多 克宏
2．発表標題 状態経験数の近似を併用した深層強化学習
3．学会等名 第37回ファジィシステムシンポジウム
4．発表年 2021年

１．発表者名
宮平 裕一，野津 亮，生方 誠希，本多 克宏

２．発表標題
JADEに対する集団外探索の追加

３．学会等名
インテリジェント・システム・シンポジウム2021

４．発表年
2021年

１．発表者名
野津 亮

２．発表標題
低コストな進化計算や強化学習のアルゴリズムの提案に向けて

３．学会等名
インテリジェント・システム・シンポジウム2021（招待講演）

４．発表年
2021年

１．発表者名
安永 恭平，野津 亮，生方 誠希，本多 克宏

２．発表標題
深層強化学習のための状態類似度の事前学習についての一考察

３．学会等名
第64回システム制御情報学会研究発表講演会

４．発表年
2020年

１．発表者名
鍔本 純也，野津 亮，生方 誠希，本多 克宏

２．発表標題
ランダムネス適応型差分進化の提案

３．学会等名
第36回ファジィシステムシンポジウム

４．発表年
2020年

１．発表者名
鍔本 純也，野津 亮，生方 誠希，本多 克宏

２．発表標題
区間を考慮した差分進化から他手法への探索途中での変更

３．学会等名
第63回システム制御情報学会研究発表講演会

４．発表年
2019年

１．発表者名
野津 亮，鍔本 純也，生方 誠希，本多 克宏

２．発表標題
差分進化における探索点群の広がりとアルゴリズムの切り替え

３．学会等名
第35回ファジィシステムシンポジウム

４．発表年
2019年

１．発表者名
野津 亮，安田 功嗣，安永 恭平

２．発表標題
深層強化学習のための状態類似性のシンプルな事前学習

３．学会等名
第29回インテリジェント・システム・シンポジウム

４．発表年
2019年

１．発表者名
安田 功嗣，野津 亮，生方 誠希，本多 克宏

２．発表標題
成長型自己組織化マップによる強化学習システムについての考察

３．学会等名
第62回システム制御情報学会研究発表講演会

４．発表年
2018年

１．発表者名 野津 亮，榊原 雅也，生方 誠希，本多 克宏
２．発表標題 差分進化における信頼区間を考慮した解候補の設定
３．学会等名 第34回ファジィシステムシンポジウム
４．発表年 2018年

１．発表者名 野津 亮，安田 功嗣，生方 誠希，本多 克宏
２．発表標題 強化学習システムにおける学習周期の無作為抽出による適応
３．学会等名 第28回インテリジェント・システム・シンポジウム
４．発表年 2018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
研究分担者	生方　誠希 (Ubukata Seiki) (10755698)	大阪公立大学・大学院情報学研究科　・准教授 (24405)	
研究分担者	本多　克宏 (Honda Katsuhiro) (80332964)	大阪公立大学・大学院情報学研究科　・教授 (24405)	

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