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　２ 年　 ６ 月　 ４ 日現在

機関番号：１４６０３
研究種目：若手研究(B)
研究期間：2017～2019
課題番号：１７Ｋ１２７５９

研究課題名（和文）階層型多目的強化学習を用いた脚ロボットの歩容自律生成

研究課題名（英文）Locomotion Control for Legged Robot using Hierarchical Multi-Objective Reinforcement Learning

研究代表者

小林　泰介（Kobayashi, Taisuke）

奈良先端科学技術大学院大学・先端科学技術研究科・助教

研究者番号：１０７９６４５２

交付決定額（研究期間全体）：（直接経費）　　3,100,000円

研究成果の概要（和文）：本研究は，脚ロボットの歩容運動を階層的な多目的最適化問題として捉えた，歩容の自律学習を目的とした．この技術の確立により，物理的な制約やトレードオフを陽に考慮可能となり，生物の自然な歩容生成が期待できる．
この課題に関連する３つの成果，(1) 継続的に学習結果を蓄積していくことが可能な正則化技術，(2) 大域的な最適解を発見可能な探索力を持つ方策，(3) 知識のモジュール化・階層化を促すニューラルネットワーク，を実現した．これらの技術を４脚ロボットの歩容を下位から上位階層モジュールに分けて順次学習するカリキュラム上で組み合わせ，製作した４脚ロボットのシミュレーションモデルでの歩容生成に成功した．

研究成果の学術的意義や社会的意義
本研究の脚ロボットの学習制御は，近年の入出力関係を直接学習してしまう大雑把なやり方では隠蔽されてしまう知識の階層関係や構成要素を，これまでの歩容に関する研究を踏まえて明示的に与えて継続的に学習を積み重ねていくことが可能な枠組みを提供しており，機械学習分野とロボティクス分野の融合領域として高い学術的意義がある．また，安定かつ高効率な歩容制御の確立はロボットの移動範囲を格段に広げて日常的にロボットが活躍するための基礎技術となり，今後のロボット共生社会の実現に繋がるものと期待できる．

研究成果の概要（英文）：The purpose of this study is to achieve locomotion control of legged robot as a hierarchical multi-objective optimization problem. With the establishment of this technology, physical constraints and trade-offs can be explicitly considered, and animal-like natural locomotion can be expected.
Three outcomes related to this study, (1) regularization technique that can continuously accumulate learning results without forgetting, (2) policy with search ability to discover global optimal solutions, (3) structured neural networks that facilitate modularization and hierarchy of knowledge. By combining these techniques on a curriculum that sequentially learns locomotion of a quadrupedal robot from the lower hierarchical modules to the upper hierarchical modules, we succeeded in generating the locomotion on a simulation model of the developed quadrupedal robot.

研究分野：知能ロボティクス

キーワード：知能ロボティクス　強化学習　多目的最適化　継続学習　歩行

様　式　Ｃ－１９、Ｆ－１９－１、Ｚ－１９（共通）

１．研究開始当初の背景

脚ロボットの歩容運動を制御するための現在の主流とである方法論では，多大な労力を払ったモデリングを必要としており，そのモデル精度により歩容の性能限界が生まれてしまっている．一方で，近年注目を集めているデータからモデルを介さずに制御則を学習する深層強化学習は，対象とする問題によっては人をも凌駕する性能を発揮し始めており，多くの注目を集めている．ただし，学習には多くの時間を要する問題や，学習結果が生物らしさを感じない不自然な歩容となりがちである．

学習した歩容が不自然になるのは，事前に問題に関する知識を敢えて一切取り入れずに入出力関係を学習する，End-to-End 学習の弊害と考えた．歩容であれば，運動方程式レベルの詳細なモデルとまではいかずとも，定性的な特性として，例えば脚が接地しているか否かによって振る舞いや制御目的が大きく変化することが考えられる．こういった定性的な事前知識を最大限に活用した学習アルゴリズムを階層的に構築することで，モデルベース制御よりも優れ，かつ生物らしさを感じさせる自然な歩容を発現できると着想に至った．

２．研究の目的

本研究の目的は，脚ロボットの歩容運動を階層的な多目的最適化問題として捉えた，歩容の自律生成手法としての階層型多目的強化学習を確立することである．その要素技術として，(1) 複数の最適化目的を学習・達成するための強化学習技術や，(2) 脚ロボットに適した問題の階層化と層ごとに学習を進めるためのカリキュラムを設計・開発する．これらの要素技術から構築される提案手法により，脚ロボットが安定かつ高効率な歩容を自律的に獲得できるようにし，その活躍の場を拡大することを目指す．

３．研究の方法

主に以下の３つの課題解決に取り組んだ．

(1) 継続的に学習結果を蓄積していくことが可能な学習正則化技術の開発
複数の最適化目的を達成したい場合に，それらを「同時に」学習しようとすると，全体としての目的が不明瞭になって困難となる一方，それらを「１つ１つ順番に」学習しようとすると，過去の学習結果を新しい目的の学習時に上書きしてしまう「破滅的忘却」という問題が知られている．この項では，過去の学習結果との関連が深いパラメータの上書きを抑制すると同時に，不必要なパラメータは新しい目的のために初期化するような正則化手法を構築する．特に，必要・不必要を分けるしきい値を適応的に与えるような設計を図る．この手法により，学習結果を必要最低限のパラメータで表現しつつ，それらを保持することで複数の目的を達成する制御器を獲得する．シミュレーションを経て，どの程度過去の学習結果を保持しておけるか，どの程度の目的を蓄積していくことができるかを検証する．

(2) 大域的な最適解を発見可能な優れた探索力を持つ方策の設計・解析
複数の最適化目的を同時に達成したい場合，全体として最も良い大域的最適解以外にも，例えば目的の１つだけを満たすといった，大域的最適解を妨げうる局所最適解が生まれてしまう．このような局所最適解（群）に妨げられずに大域的最適解へ到達するためには，強化学習の構成要素の１つである確率的な方策（制御器）が多種多様な行動を生成しつつ慎重に学習していく必要がある．この項では，所望の性質を持つものとして，スチューデントの t 分布（以下，t 分布）で確率モデル化された方策を設計する．t 分布はべき乗則に従う確率分布の裾の重さから，外れ値に頑健なモデルとして広く用いられている上に，多種多様な行動を生成する生物の優れた餌の探索能力のモデルとしても広く知られているため，これらの特性が強化学習の方策として適切に機能するかを解析する．シミュレーションにおいて目的の達成度を一般的に用いられる方策と比較する．

(3) 知識のモジュール化・階層化を容易にするニューラルネットワークの設計
異なる下位目的を適切に切り替えながら上位の目的を達成するためには，各目的を達成可能な方策モジュール群を構築することが望ましい．ただし，完全に分離したモジュールに知識を蓄積してしまうと，複数の目的間に潜む共通の部分目的を共有できずに都度再学習する必要があるため無駄が多い．この項では，単一のニューラルネットワーク内の結合の仕方をモジュール性を持つネットワークモデルを基に設計する．また，モジュールが達成目的と１対１対応するように，モジュールへの入力を与えられた目的に応じてコントロールするゲート機構を加える．これらの設計は，下位モジュール群から順番に学習を進めて上位モジュール群へと至るカリキュラムを容易に組むことを可能にする．(1)，(2) と併用した動力学シミュレータ上での脚ロボットの歩容学習実験を行い，End-to-End 学習では困難な問題であっても容易に学習可能なことを示す．

４．研究成果

(1) に関する研究成果

過去の学習結果に対するパラメータの重要度はフィッシャーの情報量として表せることが先行研究にて提案されている．パラメータが過去の学習結果を最もよく表せるであろう最適値へと拘束する強さをこの重要度に応じて決定することで，重要なパラメータを保持できる．一方で，不必要なパラメータの初期化には，重要度の逆数相当に応じた L1 スパース正則化が効果的である．この２種類の正則化を連続的に統合する手法を開発した．これらの繋ぎ目として，パラメータの最適値から重要度に応じた距離が離れるとスパース正則化の効能が現れるように設計した．しかし，前述の手法は実用的な計算コストとは言えなかったため，瞬間的なパラメータの重要度を過小評価して蓄積していくことで得られる全体の重要度を用いて２種類を単純に重み付けした手法を提案した．これらは複数の目的を設定できる強化学習シミュレーションや手書き数字の分類を２種類ずつ段階的に学習する問題で検討した．図１は後者の結果であり，正則化を加えないと過去の学習結果を保持できず (Baseline)，スパース正則化を加えずに新しい目的の学習余地が少ない従来手法（EWC）と比べて提案手法（EWCS）が新しい内容を学習・蓄積できることがわかった．

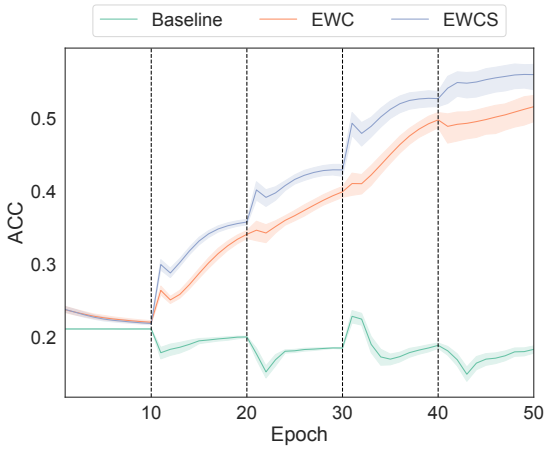

図 1 手書き数字分類の逐次学習結果

(2) に関する研究成果

前述の通り，t 分布は裾の重い分布とされているが，３種類のパラメータで定義されるモデルであり，その内の自由度パラメータによって裾の重さが決定される．そこで，自由度パラメータが無限の場合（ガウス分布）と１の場合（コーシー分布）に従って確率的に行動するエージェントで探索移動を行ったところ，明らかにコーシー分布が広範囲の探索が可能だった．また，自由度パラメータに応じた強化学習における方策改善のための勾配を解析したところ，勾配が極端な行動（外れ値）に対して定数へと収束し，保守的な更新が行われることがわかった．加えて，多変量・対角な分布を仮定すると，従来のガウス分布では各変量の更新が独立となるが，t 分布では多変量の影響を受けてどれか１種類の行動でも極端であれば更新が保守的となることがわかった．提案手法は４種類のロボット制御シミュレーション（図２参照）のどれでも，局所最適解を回避して従来手法以上の性能を獲得できることを確認した．

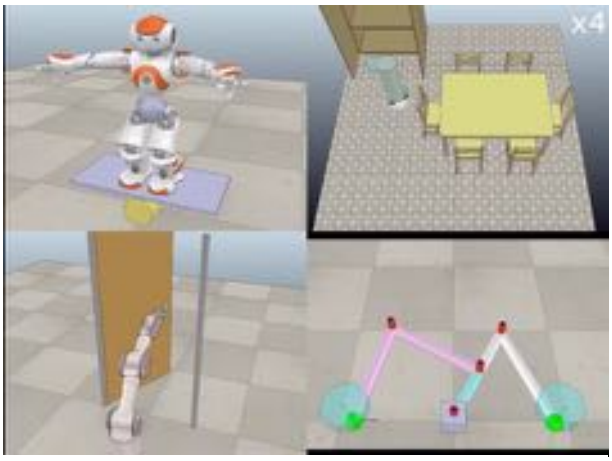

図 2 局所最適解を含むロボット制御問題

(3) に関する研究成果

モジュール性を有するネットワークモデルとして，複雑ネットワークの一種であるフラクタルネットワークを採用した．この結合の仕方に従って，ニューラルネットワークの一種であるリザーバコンピューティングの中間層を事前設計することで，モジュール単位での活性化を促しやすく，かつ隣接モジュールとニューロンを共有することで共通の部分目的を扱いやすい設計を得た．これに合わせて，モジュール中心であるほど目的に特化し，外縁では共有化しやすいような入力ゲート機構を設けた．この設計されたネットワークを最大限に活用するように，２階層の歩容学習アルゴリズム（図３参照）を構築した．

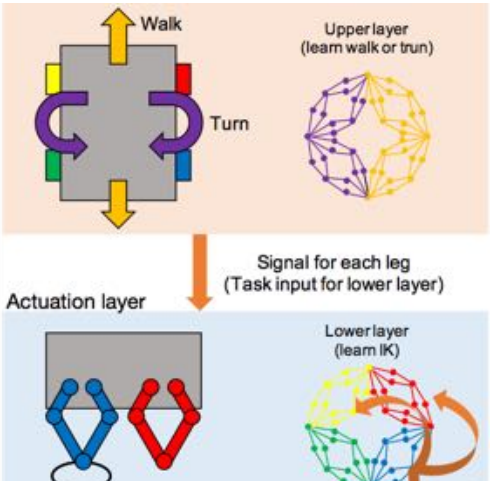

図 3 ２階層の歩容学習アルゴリズム

すなわち，下位階層のモジュールから(1)，(2)を併用しつつ順番に学習させ，かつ上位階層のモジュールを学習する際に下位モジュールを再調整するようなカリキュラムを組んだ．また，下位モジュールは各脚に対応するため，単脚で脚の動かし方を学習させた後に，その学習結果を他モジュールへと転写する方法も考案した．図４に示される開発した４脚ロボットに対応するシミュレーションモデルを用いて前進・旋回タスクの学習をカリキュラムに沿ったところ，End-to-End 学習では学習に失敗して歩容を獲得できなかった一方で，提案手法は前進・旋回ともに効率良く学習できることを確認した．

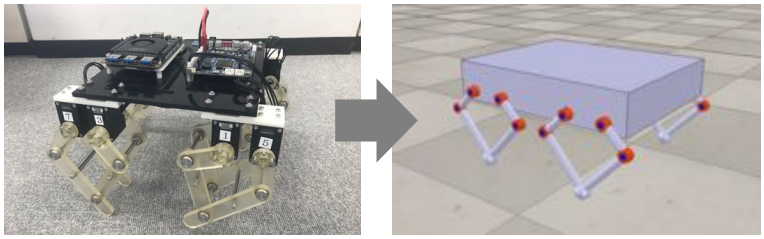

図 4 開発した４脚ロボットとシミュレーションモデル

５．主な発表論文等

〔雑誌論文〕　計1件（うち査読付論文　1件／うち国際共著　0件／うちオープンアクセス　0件）

１．著者名 Taisuke Kobayashi	４．巻 49
２．論文標題 Student-t policy in reinforcement learning to acquire global optimum of robot control	５．発行年 2019年
３．雑誌名 Applied Intelligence	６．最初と最後の頁 4335-4347
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/s10489-019-01510-8	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7件（うち招待講演　0件／うち国際学会　3件）

１．発表者名 Taisuke Kobayashi and Toshiki Sugino
２．発表標題 Continual Learning Exploiting Structure of Fractal Reservoir Computing
３．学会等名 International Conference on Artificial Neural Networks（国際学会）
４．発表年 2019年

１．発表者名 杉野 峻生，小林 泰介，杉本 謙二
２．発表標題 フラクタルリザーバコンピューティングを用いた４脚ロボットの階層強化学習
３．学会等名 ロボティクス・メカトロニクス講演会
４．発表年 2019年

１．発表者名 小林 泰介
２．発表標題 パラメータの定着とスパース化を統合した正則化による継続学習
３．学会等名 日本ロボット学会学術講演会
４．発表年 2019年

１．発表者名
杉野 峻生, 小林 泰介, 杉本 謙二

２．発表標題
フラクタルリザーバコンピューティングを用いた継続学習

３．学会等名
ロボティクス・メカトロニクス講演会

４．発表年
2018年

１．発表者名
Toshiki Sugino, Taisuke Kobayashi, Kenji Sugimoto

２．発表標題
Continual Learning using Modularity of Structured Reservoir Computing

３．学会等名
SICE Annual Conference（国際学会）

４．発表年
2018年

１．発表者名
Taisuke Kobayashi

２．発表標題
Check Regularization: Combining Modularity and Elasticity for Memory Consolidation

３．学会等名
International Conference on Artificial Neural Networks（国際学会）

４．発表年
2018年

１．発表者名
小林 泰介

２．発表標題
大域的最適解を目指すActor-Critic強化学習

３．学会等名
日本ロボット学会学術講演会

４．発表年
2017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