The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

SUNA によるシミュレーション上での二足歩行動作学習

Learning to Bipedal Walking on Simulation with SUNA

井上　湧太　　ダニロ・ヴァスコンセロス・ヴァルガス
Yuta Inoue　　Danilo Vasconcellos Vargas

九州大学
Kyushu University

SUNA is currently one of the most adaptive neuroevolution methods which is able to tackle different problems efficiently. However, many questions remain unanswered. In this research, we applied SUNA to the bipedal-walking problem and evaluate it general learning properties. The results show that even without any modificiations SUNA is able to learn in this environment. Moreover, contrary to many other methods, it is continuously improving its average rewards showing a near open-ended learning.

1. はじめに

Spectrum-diverse Unified Neuroevolution Architecture (SUNA)[Vargas 17] は、異なる問題を同一の手法で効率的に学習するためのニューロエボリューションの手法である。しかしながらその性質についての多くは調べられておらず、汎用性や計算速度での優位性などの性質は明らかになっていない。本研究では SUNA の汎用性について調べるため、二足歩行シミュレーションを用いて SUNA による学習を行った。

SUNA は Unified Neural Model (UNM) と呼ばれる特殊な構造を用いて汎用性を実現しており、また効率性は UNM の各ニューロンの特徴にしたがって探索を行うことで実現されている。UNM では、同じニューロエボリューションの手法である NEAT[Stanley 02] で使用されるニューロンに改良を加えた拡張ニューロンが使用される。このニューロンでは NEAT で使用されるニューロンの持つ特徴に加え、それぞれ異なる時定数を持ち、異なる活性化関数を選択できるという特徴がある。異なる活性化関数を選択できることは、単一の活性化関数を持つニューロンのみで構成されるネットワークと比較して、ある小さなネットワークを 1 つのニューロンで表すことができる可能性があるため、構造を簡単にすることに貢献する。

また UNM では、多様な問題に対応できるように、状況に応じてネットワークの構造を変えるための制御ニューロンとニューロモジュレーションを導入している。制御ニューロンは入力に応じて、制御対象のニューロンごとに重み付けされた制御信号を出力する。各ニューロンは受け取った制御信号の総和が閾値以上であればそれぞれのニューロンの入力に従って出力を行い、そうでなければ出力を行わない。まだニューロモジュレーションはあるニューロンの内部状態に応じて結合強度を変化させることでネットワークの構造を変化させる。

以上で上げたように UNM は高い自由度を持つため、その探索空間は膨大なものになる。これを効率的に探索するため、Danilo & Murata[Vargas 17] ではネットワークの含むニューロンの特徴によって個体を種別し、それぞれの種の中で最も高い適応度を持つものを変異させたものについて探索を行った。その世代のすべての個体を Novelty Map を用いて含まれる個体同士の距離が最も大きくなるように選抜し、選抜された個体を中心として下位個体群とする。それぞれの下位個体群のうち最も適応度の高い個体と、それらを変異させた個体を次の世代として評価を行う。これを行うことで探索の方向を抽出した特徴量に応じて分散することができるため、効率的な探索が可能になると考えられる。

Contact: 井上 湧太、九州大学、福岡県福岡市西区周船寺 1 丁目 9 番地 19 パークホームズ周船寺 901 号、yi1306c12@gmail.com

図 1　OpenAI Gym[Brockman 16] の二足歩行シミュレーション（ BipedalWalker-v2）

2. 方法

学習環境には OpenAI Gym の 2 次元二足歩行シミュレーション環境である BipedalWalker-v2（ 図 1）を用いた [*1]。入力は、すべての入力を使用できる場合と、使用できる入力が制限される場合に分けて実験を行った。すべての入力が使用できる場合、使用可能な入力は胴体（ 図 1 中の紫色の多角形）の進行方向に対する速度、上下方向に対する速度、角度、角速度、両足の股関節、膝関節の角度、角速度、両足それぞれが地面と接しているかどうかなどの 24 つの入力を与えた。また入力は両足それぞれの股関節、膝関節の角度、それぞれの足が地面に付いているかどうかの計 6 つとした。出力は両足の股関節、膝関節の角速度の計 4 つであった。報酬はロボットの前進に応じて与えられた。また胴体の傾きや出力の大きさによって負の報酬が与えられた。到達地点まで達した場合 +300、ロボットの胴体（ 図 1 中の紫色の多角形）が地面（ 図 1 中の下部の緑色の部分）に接触した場合-100 の報酬を与え、試行を終了した。前述した条件で終了しなかった場合は、1 試行で最大 300 ステップを行った。各個体に対して 5 回の試行を行い、試行ごとの報酬の平均を適応度とした。SUNA のパラメータは Danilo & Murata

*1　https://gym.openai.com/envs/BipedalWalker-v2

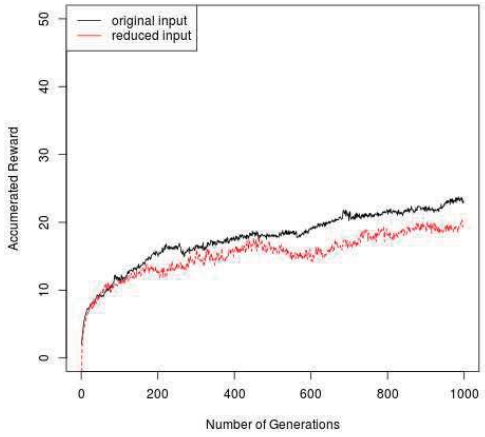

図 2　各世代で最も高かった適応度の、30 回の平均

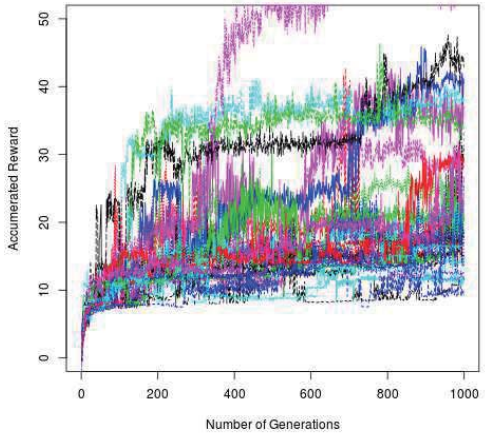

図 3　すべての入力を使用可能にした場合の各世代で最も高かった適応度の、実験 30 回のそれぞれの遷移

(2017)[Vargas 17] と同じものを使用した。1 世代あたりの個体数は 100 とし、1000 世代分の学習を行った。

3.　結果

実験を 30 回行った場合の、それぞれの学習において各世代での最も高い適応度の平均を図 2 に示す。また以下に 30 回行った実験それぞれでの、各世代での最も高い適応度の世代ごとの遷移を図 3 と図 4 に示す。

4.　考察

表 1 に、本実験の結果と同一の環境を用いて実験を行った [Zhang 17] の結果を示す。ただし SUNA の結果は 10^7 ステップの結果と合わせるため、100 世代目前後のものを使用している。ステップ数は正確に対応していないため直接の比較はできないが、SUNA の結果は NEAT と同程度であった。

入力数が 24 の場合と 6 の場合では、24 の場合のほうが良い結果を出している（ 図 2）。入力を限定した場合では実験ごとの

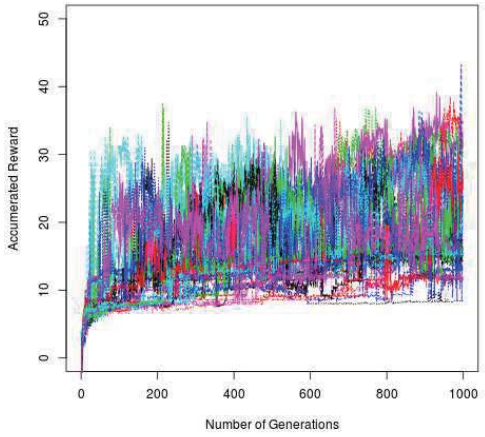

図 4　入力を制限した場合に各世代で最も高かった適応度の、実験 30 回のそれぞれの遷移

表 1　SUNA とその他の手法の結果の比較。Shangtong & Osmar（ 2017）[Zhang 17] の結果を使用した。ただしすべての値はグラフ（ Shangtong & Osmar（ 2017）,Figure. 2）から読み取った値である。

methods	SUNA	NEAT	NES	CMA	P3O
rewards	10	10	50	-100	300

適応度が世代数の増加とともに安定して上がらず（ 図 4）、入力数が多いほうが学習はスムーズであった（ 図 3）。SUNA では突然変異の際、ニューロン間の結合強度を変化させるかどうか決めてから、変化させるのであればそれぞれの結合強度を一様分布で得られた値を用いて変化させるという手段を取る。この方法では、これまでの進化で得られた有用なネットワークを保持しながら新しい構造を探索するようなことは難しいため、予め必要な情報が入力として用意されている場合のほうが学習に有利であると考える。

参考文献

[Brockman 16] Brockman, G., Cheung, V., Pettersson, L., Schneider, J., Schulman, J., Tang, J., and Zaremba, W.: OpenAI Gym, *CoRR*, Vol. abs/1606.01540, (2016)

[Stanley 02] Stanley, K. O. and Miikkulainen, R.: Evolving Neural Networks Through Augmenting Topologies, *Evol. Comput.*, Vol. 10, No. 2, pp. 99–127 (2002)

[Vargas 17] Vargas, D. V. and Murata, J.: Spectrum-Diverse Neuroevolution With Unified Neural Models, *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–15 (2017)

[Zhang 17] Zhang, S. and Zaiane, O. R.: Comparing Deep Reinforcement Learning and Evolutionary Methods in Continuous Control, *arXiv* (2017)