動的環境下でのエージェント行動選択パラメータ自動学習法の提案

Proposal for automatic learning of agent action selection parameters in dynamic environments

下川　大樹[1*]　吉田　直人[2]　栗原　聡[1]
Daiki SHIMOKAWA[1] and Naoto YOSHIDA[2] and Satoshi KURIHARA[1]

[1] 慶應義塾大学理工学部
[1] Faculty of Science and Technology, Keio University
[2] 慶應義塾大学大学院理工学研究科
[2] Graduate School of Science and Technology, Keio University

Abstract: 現在，特化型人工知能の普及は進んでいるが，汎用型人工知能の実現には至っていない．汎用型人工知能の実現には動的で複雑な環境下で行動選択ができる手法が必要であり，そのための手法として Agent Network Architecture(以下 ANA) のような手法が過去にも提案されている．しかしながら，ANA を動作させるためのパラメータは膨大で尚且つ手動で決めていた．本論文では進化計算手法を使用し，ANA のパラメータ調整を自動的に行う方法を提案する．結果は，差分進化法を使用した場合が高い適応度を獲得することがわかった．

1　はじめに

現在，限定された領域や環境に特化して学習や処理を行う特化型人工知能は日常生活に当たり前のように登場するようになってきたが，自分で目的を持って行動を行い，人間に近い問題処理能力を持つ汎用型人工知能の実現には至っていない．

一方，人間は反射的な処理と論理的な思考処理の両方が行われている [1]．これらの処理を実現するプランニング手法として，エージェント同士がネットワークで接続され，活性伝搬により適切なエージェントが発火することで適切な行動系列を求める手法があり，Agent Network Architecture（以下 ANA）[2] や，著者らにて MRR-planning[3] や，その拡張で [4] などをこれまで提案している．これらの手法の特徴は，即応的な処理と熟考的な処理の両方を行うことができるところにある．その反面，このエージェント活性伝搬ネットワークの構成や活性伝搬させる度合いの調整は人手で行う必要があり，そのために大規模化や汎用性を持たせにくいという問題点があった．よって，これらを自動で調整できれば，このエージェント活性伝搬ネットワークを特定の環境のみでなく汎用的な環境下でも動作することが可能になると考えられる．

そこで，本研究では非勾配法の進化計算手法を用いて，エージェント活性伝搬ネットワークのパラメータ調整の自動化する手法の提案した．そして進化計算手法である差分進化法を用いた場合が他の進化計算手法より高い適応度を獲得できることを確認できた．

*連絡先：慶應義塾大学理工学部
〒 223-8522 神奈川県横浜市港北区日吉 3-14-1
E-mail: jump1204@keio.jp

2　先行研究

2.1　プランニング手法の先行研究

2.1.1　ANA

ANA[2] とは 1991 年に Pattie Maes によって提案された行動ベースのアーキテクチャで，簡単な機能を持つ多数の分散したモジュールの集合のことである．エージェント活性伝搬ネットワークは図 1 のような構造となっている．

各モジュールは活性レベル，閾値，add list，delete list，condition list を持つ．これらの list を通じてそれぞれのモジュール同士は行動の整合性を持って繋がっている．また，各モジュールの活性レベルは隣接しているモジュールと現在の環境から刺激を蓄積した値のことを指し，活性レベルが閾値を超えるとそのモジュールは実行される．

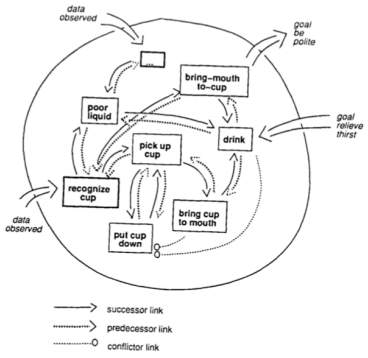

図 1: ANA のネットワーク構造 [2]

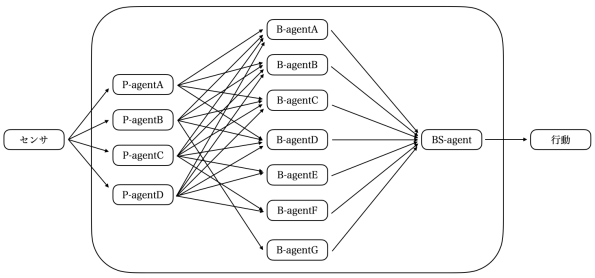

図 2: MRR-planning のエージェントネットワーク構造

2.1.2 MRR-planning

MRR-planning の構成は図 2 のようになっている．

MRR-planning は P-agent(プランニングエージェント)，B-agent(行動エージェント)，BS-agent(行動選択エージェント) の 3 つのエージェント群からなるエージェント活性伝搬ネットワークである．P-agent はそれぞれのエージェントが目的を持っており，B-agent はそれぞれのエージェントが行動を持っており，BS-agent は B-agent の中で最も優先度が高い行動を選択する．

行動決定までの流れは，P-agent がセンサから情報を受け取って，その情報を元に B-agent に活性値を渡す．B-agent はその活性値を蓄積し，その蓄積した活性値がある閾値を超えたら，その B-agent が BS-agent により選択され，B-agent の内容が実行される．この情報の受け渡し方により即応的な行動と熟考的な行動を使い分けることができる．

2.1.3 本研究で提案するエージェントネットワーク

Yoshida ら [4] は Tile World 上で動くエージェントに ANA と MRR-planning をベースとした図 3 のエージェント活性伝搬ネットワークを組み込んだ．

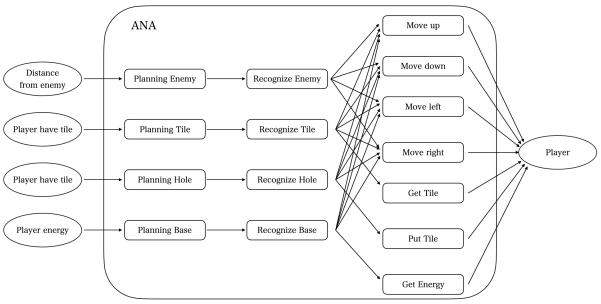

図 3: 本研究におけるエージェントネットワークの構造

Planning モジュールでは環境からの情報を利用して，どの物体に対してプランニングを行うかを決定する．Recognize モジュールでは Planning モジュールによって決定された物体がエージェントから見てどの方向にいるかを探す．Action モジュールでは Recognize モジュールの結果からどの行動をするかを決めるモジュールである．このエージェント活性伝搬ネットワークは ANA と同様に，各モジュールが閾値と他のモジュールへの刺激の強さを決める伝播値を持ち，本研究では各モジュールの持つ閾値と伝播値のことをパラメータと呼ぶ．

このエージェント活性伝搬ネットワークでは，ANA と MRR-planning の利点である即応的な行動と熟考的な行動の両方を兼ね備えており，整合性の保たれた行動の中で創発的な行動が起こる．しかし，このネットワークの活性伝播の度合いは手動で決めていたので，それが最適解とは限らない．そのため，本研究ではこのエージェント活性伝搬ネットワークのモジュールの持つ活性伝播に関わるパラメータを 2.2 節で述べる提案手法を用いて自動調節する．

2.2 パラメータ学習手法

本節では ANA でパラメータ更新を行うために使用した進化計算手法の解説を行う．進化計算手法とは生物の進化を模倣した手法で，より環境に適応できた個体を残すことで近似解を探索する最適化手法である．

エージェント活性伝搬ネットワークを持つエージェントの報酬はプランニングを行った結果として得られた環境の変化やエージェントの状態から得られるため報酬関数は非勾配になる．したがって，パラメータ調整は勾配法の最適化手法は使用できない．また，パラメータとして閾値と刺激の強さがあることでエージェント活性伝搬ネットワークは即応的な行動と熟考的な行動の両方を行うことができる．強化学習を用いた調整では閾値や伝播値を一度 Q 値等に置き換えてパラメータの更新をする必要があり，情報の損失が起こる可能性

が考えられる．そのため本研究では上記を満たす進化計算手法を用いてパラメータの調整を行った．

本研究では進化計算手法である遺伝的アルゴリズム [5] と粒子群最適化法 [6]，差分進化法 [7] を適用し，それぞれの適応度による比較を行った．

3　提案手法

本節ではエージェント活性伝搬ネットワークに進化計算手法を適用して，パラメータを自動調整する手法について説明する．

3.1　進化計算手法のエージェント活性伝搬ネットワークへの適用

進化計算手法をエージェント活性伝搬ネットワークへ適用する手法の解説する．

まず，初期個体の生成のフェーズである．本研究ではエージェント活性伝搬ネットワークのパラメータとして，各モジュールの閾値と伝播値の 2 つを扱う．そのため，閾値を T，伝播値を I，モジュール数を N 個とすると，1 個体は N 個の閾値と N 個の伝播値を持ち，それぞれをランダムで $T_{min} \leq T \leq T_{max}$，$I_{min} \leq I \leq I_{max}$ の範囲で与える．

次に，初期個体の生成フェーズで生成した個体の適応度を計算する．適応度関数を f とすると各個体の適応度は式 1 と表せる．

$$fitness = f(T_1, T_2, , , T_N, I_1, I_2, , , I_N) \qquad (1)$$

その後，次世代の個体を生成する．次世代の個体生成では遺伝的アルゴリズムや粒子群最適化法，差分進化法のそれぞれの使用する進化計算手法を用いて個体の更新を行う．

この適応度計算→個体更新の繰り返しを終了条件を満たすまで続け，終了条件を満たした場合は最も適応度の高い個体を出力してパラメータの学習を終了する．

4　実験環境

4.1　Tile World 概要

本節では本研究の実験で使用した Tile World[8] についての説明を行う．従来，Tile World は様々なプランニングの研究の動的な実験環境として使用されており，本研究でも同様に Tile World を提案手法の実験環境として使用した．

本研究での Tile World の初期状態は図 4 である．Tile World にはエージェント，タイル，ホール，敵エージェント，エネルギー供給ベース，障害物が存在する．エージェントは効率よくタイルとホールを集めながら，敵が近づいてきたら逃げ，エネルギーが少なくなるとエネルギー供給ベースでエネルギーを補給するという多目的な処理を必要とする．

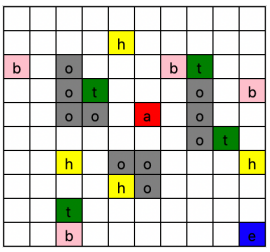

図 4: 本研究における Tile World の初期状態

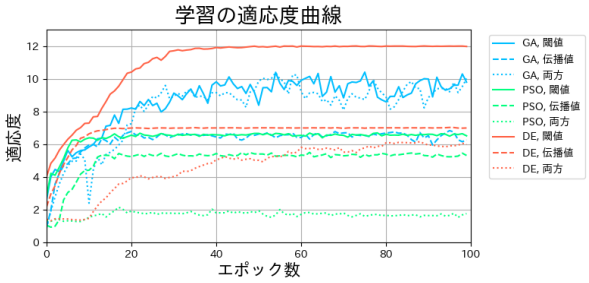

図 5: 9 パターンの学習方法による適応度曲線

4.2　学習方法による適応度の比較実験

本実験は学習方法による適応度の比較を図 4 上で行った．ここで学習方法とは，学習対象であるエージェント活性伝搬ネットワークのパラメータの伝播値を固定して閾値のみを学習させる場合，閾値を固定して伝播値のみを学習させる場合，両方を学習させる場合の 3 パターンと，学習手法である遺伝アルゴリズム，粒子群最適化法，差分進化法の 3 パターンの計 9 パターン学習をした．そのため，それぞれの学習方法における学習の適応度の変化を比較した．

5　実験結果・考察

9 パターンの学習方法による適応度曲線を図 5 に示す．

5.1　進化計算手法による比較

遺伝的アルゴリズムによる学習は高い個体平均適応度となったが，適応度の分散は非常に高い値となって

いる．これは突然変異が起きるとそれまで良い適応度を導出するパラメータがランダムな値に変わってしまうことによって，収束の安定性が無くなったことが原因と考えられる．粒子群最適化法による学習は個体平均適応度の平均が低く，局所解に収まってしまった．また，差分進化法による学習は個体平均適応度が高く，分散が小さくなり，局所解に収まらなかったことが確認できた．これは差分進化法は個体を生成する際に 4 個体のパラメータを使用するため，母集団の多様性が高くなることが原因と考えられる．

5.2 学習対象パラメータによる比較

閾値のみの学習は他のパラメータの学習と比較して最も高い適応度を獲得できた．反対に伝播値のみの学習と閾値，伝播値両方の学習の適応度は局所解に収束してしまった．伝播値のみの学習において局所解に収束した原因として，今回のエージェント活性伝搬ネットワークにおいて他のモジュールに与えられる刺激の強さをモジュールの伝播値と関数を使用して決めていることが挙げられる．そのため伝播値を変更させると関数による出力値が大きく変わってしまいネットワークの安定性がなくなることが考えられる．また，上記でも説明したように伝播値の学習は安定性がなかったため，その影響で閾値と伝播値の両方を学習は安定性がなかったと考えられる．よって，閾値のみの学習がより安定して学習を進めることができたと考えられる．

まとめると，進化計算手法として差分進化法を用いて，伝播値を固定し，閾値のみを学習させる場合がより安定して高い適応度を獲得できることがわかる．

6 結論

本研究ではエージェント活性伝搬ネットワークのパラメータの調整の自動化における提案と Tile World を用いた評価実験を行った．本実験においては進化計算手法である差分進化法を用いて，エージェント活性伝搬ネットワークの伝播値を固定して閾値のみを学習させる方法が現在の環境に合わせた効率的な行動を選択することができた．つまり，進化計算手法でエージェント活性伝搬ネットワークのパラメータ調整を自動調整することが可能であることを確かめられた．

本研究で確認できていないことは 2 つある．

1 つ目は元あるパラメータを使用し，再学習する方法である．拡張した機能や行動があった際に拡張したモジュールや元からあるモジュールのパラメータを調整する必要がある．

2 つ目はモジュール数が膨大になった際の挙動についてである．現実世界で本研究で用いたエージェント活性伝搬ネットワークでプランニングを行う際には，モジュールの数は膨大になることが予想させる．そのため，モジュールの数が膨大になった際にいかにして進化計算手法を用いてパラメータを獲得するかが問題となる．

謝辞

本研究は，NEDO・人と共に進化する次世代人工知能に関する技術開発事業「インタラクティブなストーリー型コンテンツ創作支援基盤の開発」の助成を受けたものである．

参考文献

[1] Dennis, L. A. and Fisher, M.: Verifiable Self-Aware Agent-Based Autonomous Systems, *Proceedings of the IEEE* (2020).

[2] Maes, P.: The agent network architecture (ANA), *Acm sigart bulletin*, Vol. 2, No. 4, pp. 115–120 (1991).

[3] Kurihara, S., Aoyagi, S. and Onai, R.: Adaptive selection of reactive/deliberate planning for the dynamic environment, *European Workshop on Modelling Autonomous Agents in a Multi-Agent World*, pp. 112–127 (1997).

[4] 吉田直人，高屋英知，加藤慶彦， 栗原聡ほか ：動的環境におけるマルチエージェントプランニングの提案，研究報告知能システム (ICS)， Vol. 2020, No. 9, pp. 1–7 (2020).

[5] Holland, J.: Adaptation in natural and artificial systems: an introductory analysis with application to biology, *Control and artificial intelligence* (1975).

[6] Kennedy, J. and Eberhart, R.: article swarm optimization, *International Conference on Neural Network*, Vol. 4, pp. 1942–1948 (1995).

[7] Storn, R. and Price, K.: Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces, *Journal of global optimization*, Vol. 11, No. 4, pp. 341–359 (1997).

[8] Pollack, M. E. and Ringuette, M.: Introducing the Tileworld: Experimentally evaluating agent architectures, *Association for the Advancement of Artificial Intelligence*, Vol. 90, pp. 183–189 (1990).