

# 博士論文審査結果報告書

## 論 文 題 目

Study on Robustness and Adaptability of  
Genetic Network Programming with  
Reinforcement Learning for Mobile Robot

申 請 者  
Siti SENDARI

情報生産システム工学専攻  
ニューロコンピューティング研究

2013 年 2 月

有向グラフ構造をもつ Genetic Network Programming(GNP)は、2000年に平澤が提案した進化論的アルゴリズムであり、①部分観測マルコフプロセスを容易に実現できる、②判定ノードおよび処理ノードの重複利用が可能なたためコンパクトな解を生成できる、③ノード遷移を活用するためメモリ機構を有しているなどの特徴をもっている。また、世代ごとの進化と個体の学習を同時に行うことにより GNP と Reinforcement Learning(RL)を統合した GNP-RL の研究もおこなわれている。しかし、GNP-RL の研究は主に GNP の訓練時の進化性能とテスト時の汎化性能の向上のためであり、GNP-RL により制御されるシステムの頑健性と適応性の向上を目指したものではなかった。

本論文は、GNP-RL のモデルの拡張に関するものであり、GNP-RL により制御されるシステムの頑健性の向上(予測される環境変動に対応できるシステムを構築すること)と適応性の向上(環境変動に適応して変化していくシステムを構築すること)を実現する各種の方式を提案し、Khepera ロボットの制御の問題を使用して評価している。

具体的には、頑健性の向上に関しては、GNP-RL の訓練時に多様なノード遷移が可能な Exploration を重視した進化を行い、訓練時とは異なる将来の環境変動に強いシステムを構築する Fuzzy GNP-RL を提案し評価している。また、ロボットセンサの故障に対する適応性の向上に関しては、GNP のノード遷移の変化に柔軟性を持たせるため、GNP の連結した 2 個の判定ノードあるいは処理ノードに構造の異なる RL をそれぞれ実装し、これらが協調的に動作する 2 つの方式の Fuzzy GNP with Two Stage Reinforcement Learning(Fuzzy GNP-TSRL)を提案し評価している。さらに、RL の Exploration と Exploitation のバランスを制御する  $\epsilon$  greedy policy パラメータ と RL の速度を制御する学習パラメータ  $\alpha$  を Khepera ロボットセンサの故障の経緯とともに同時に変化させる方式を提案し評価している。

第 1 章では、GNP-RL を解説し、Fuzzy GNP-RL と Fuzzy GNP-TSRL がこれらによって制御されるシステムの頑健性と適応性の向上にとって有効であるという着想に至った経緯および期待できる効果を従来方式と比較しながら述べ、本論文の内容を要約している。

第 2 章では、GNP-RL の判定ノードにファジィメンバシップ関数を定義する Fuzzy GNP-RL を提案し評価している。これによって、判定ノードの分岐がファジィメンバシップ関数より計算される確率を使用するため頑健なロボットコントローラが設計できること、また、ロボットセンサの連続変数を閾値で区分する従来方式の Sharp Boundary Problem を回避することが可能になり GNP-RL の性能が向上することを示している。さらに、ロボットセンサの入力にノイズを加え

て Fuzzy GNP-RL を訓練し、あわせて、ファジィメンバシップ関数を進化により最適化すると更なる頑健性の向上が図れることを示している。

シミュレーションでは、壁伝い制御を行う Khepera ロボットの人工脳を Fuzzy GNP-RL と GNP-RL を使用して実現し、ロボットセンサの入力にノイズを加えた場合と加えない場合について比較評価を行っている。ロボットが壁伝いに移動した正規化距離を適合度として頑健性の評価を行った結果、ノイズ付 Fuzzy GNP-RL, ノイズ付 GNP-RL, ノイズ無 Fuzzy GNP-RL, ノイズ無 GNP-RL の適合度は、それぞれ、0.470, 0.434, 0.143, 0.401 であった。これにより、ロボットセンサの入力にノイズを入れて訓練する Fuzzy GNP-RL が頑健性の向上の点で最も優れていることを明らかにしている。

第 3 章では、GNP の連結した 2 個のノードに構造の異なる RL をそれぞれ実装し、これらが協調的に動作する Fuzzy GNP with Two Stage Reinforcement Learning(Fuzzy GNP-TSRL)を提案し評価している。2 個の RL の協調的動作とはステージ 1 の RL では環境変動への適応に限界がある場合に、ステージ 2 の RL がこれを補完してより環境変動に適応するように Fuzzy GNP-TSRL のノード遷移を変更していくことであると説明している。なお、第 3 章では、ステージ 2 の RL の行動をノードのブランチ(Branch)を他のどのノードに接続するかを選択(Selection)と定義しているため、提案方式を Fuzzy GNP-TSRL(BS)と名付けている。

シミュレーションでは、Fuzzy GNP-TSRL(BS) と Fuzzy GNP-RL を使用して実現した壁伝い制御を行う Khepera ロボットの人工脳を適応性の点で評価している。ロボット稼働後にロボットセンサに故障が発生した場合のロボット軌道の回復プロセスを、ロボットが壁伝いに移動したセンサ故障後の正規化平均距離で比較している。その結果、Fuzzy GNP-TSRL(BS) と Fuzzy GNP-RL の正規化平均距離は、それぞれ、0.120, 0.060 であった。これにより、Fuzzy GNP-TSRL(BS) の正規化平均距離が Fuzzy GNP-RL より約 50%優れていることを明らかにしている。

第 4 章では、Fuzzy GNP-TSRL(BS)を拡張した方式を提案し評価している。具体的には、Fuzzy GNP-TSRL(BS)では固定していた Exploration と Exploitation のバランスを制御する  $\epsilon$  greedy policy パラメータ と強化学習の速度を制御するパラメータ  $\alpha$  をロボットセンサの故障の経緯とともに同時に変化させる方式を提案し評価している。これはロボットセンサの故障直後は  $\epsilon$  と  $\alpha$  を大きな値に設定し、その後徐々に値を小さくしていく方式で、ロボットセンサの故障を Fuzzy GNP-TSRL(BS)の Q 値の変化により検出するのが特徴であることを示している。

シミュレーションでは、パラメータ  $\epsilon$  と  $\alpha$  を変化させる Fuzzy GNP-TSRL(BS) を使用して実現した壁伝い制御を行う Khepera ロボットの人工脳を、固定パラメ

ータの Fuzzy GNP-TSRL(BS)と適応性向上の点で比較評価している。第 3 章と同様、ロボット稼働後にロボットセンサに故障が発生した場合のロボットの正規化移動平均距離を計算している。その結果、パラメータ  $\epsilon$  と  $\alpha$  を変化させる Fuzzy GNP-TSRL(BS) の正規化移動距離は 0.170 であり、固定パラメータの Fuzzy GNP-TSRL(BS)に比較し適応性が約 42%向上することを明らかにしている。

第 5 章では、Fuzzy GNP-TSRL(BS)とは異なり、ステージ 2 の RL に Credit GNP を使用した Fuzzy GNP-TSRL を提案し評価している。Credit GNP では GNP のすべてのノードにノードの有効性を判断する新たな Credit ブランチを追加し、Sarsa Learning で学習した Credit ブランチの Q 値と従来のブランチの Q 値を比較し、その大小により Credit ブランチを選択している。ステージ 2 の RL の行動が GNP の Credit ブランチの選択(Selection)と定義しているため、提案方式を Fuzzy GNP-TSRL(CS)と名付けている。Credit ブランチを選択した場合には Credit ブランチの Q 値を更新せず、また、Credit ブランチが選択されたノードを有害であるとしてスキップしている。その結果、1 ステップの処理ノード数が増加するため、ステージ 1 の RL の環境変動への適応限界を克服する可能性が増大し、更なる適応性の向上が期待されることを説明している。

シミュレーションでは、第 4 章と同様、ロボット稼働後にロボットセンサに複数個の故障が発生した場合の正規化移動平均距離を Fuzzy GNP-TSRL(CS)と Fuzzy GNP-TSRL(BS)で比較したところ、それぞれ、0.230, 0.190 であった。これより、Fuzzy GNP-TSRL(CS)が壁伝いロボットの適応性向上の点で GNP-TSRL(BS)より約 21%優れていることを明らかにしている。

第 6 章では、本論文で提案し評価を行った Khepera ロボット制御の頑健性と適応性を向上させるための GNP の新しいアルゴリズムの研究成果を総括している。

以上、本論文では、GNP により制御されるシステムの頑健性と適応性を向上させるための新しいアルゴリズムを提案し、Khepera ロボットの制御の問題を使用してその有効性を検証している。従って、モバイルロボットの人工脳の構築における頑健性と適応性の向上に寄与するところが大きい。よって、本論文は博士(工学)の学位論文として価値あるものと認める。

2013 年 1 月 21 日

主査	早稲田大学	教授	博士(情報工学)	(九州工業大学)	古月敬之
	早稲田大学	教授	工学博士	(早稲田大学)	吉江修
	早稲田大学	教授	博士(工学)	(早稲田大学)	藤村茂
	早稲田大学	名誉教授	工学博士	(九州大学)	平澤宏太郎