

博士論文審査報告書

論文題目

自律型ロボットにおける制御回路構造の学習手法
～オンライン・リアルタイムな
ネットワーク構造の強化学習～

Learning Method for Control Circuit in
Autonomous Robot
～ Online Real-Time Structural
Reinforcement Learning～

申請者

金 天海

Chyon Hae Kim

機械工学専攻 精密工学研究

2008年 2月

宇宙・海底の未知環境を自律的に探査するロボットや、対人コミュニケーションロボット等の自律型ロボット実現への期待が高まっている。自律型ロボットには、場面に応じた多様な行動を実現するための自律性が必要であるが、そのための効果的な制御系構築方法は未だ完成していない。この自律制御系を構築する方法論としては、設計段階において固有の制御系を実装するのではなく、学習制御によって制御則自身を獲得する機能を実装することが効果的であると言われている。自律型ロボットのための学習制御器には、次の 5 要素が必要となる。すなわち、新たな行動を創発するための行動創発、獲得した知識を効率良く保持するための汎化・抽象化、タスク・環境毎の調整を避けることができる柔軟性、動的な環境に対応するための随時学習を実現するオンライン性、学習内容の忘却を抑制する漸次性である。これまでの学習制御器でこれらを全て同時に満たすものは存在していない。

本論文は、オンライン・リアルタイムなネットワーク構造の強化学習を実現できる学習制御器を開発し、その学習制御器において漸次性の確認を行うことで、上述の 5 要素を同時に実現できる枠組みを構成することを目的としている。そして具体的に独自の自己組織化回路素子 **Self-Organizing Network Elements (SONE)** を提案している。SONE はネットワーク型の回路構造を学習によって獲得することで、汎化・抽象化、柔軟性を確保することができる。また、ネットワーク内に強化信号を伝播して各素子の有効度を評価することで学習を行い、強化学習に基づいた行動創発も実現することができる。さらに、伝播した強化信号をネットワーク内の各素子へ蓄積することで、漸次性とオンライン性の高い学習も実現することができ、自律型ロボットの学習制御器の条件を満たすものである。

本論文は、以下の 6 章から構成されている。

第 1 章では、自律型ロボットの学習制御則を設計者が予め作成することの限界と学習制御の必要性を、未知環境探査ロボットやコミュニケーションロボットの例をとり説明している。

第 2 章では、自己組織化回路素子 **SONE** を提案している。また、**SONE** による具体的な学習制御器の構成例として、自己組織化論理回路の構成法を述べ、自己組織化論理回路に関する数学的な証明について述べている。**SONE** では、ネットワークを構成するために必要となる素子のそれぞれを独立したシステムとして記述している。また、各素子が持つ機能として、次の 3 つの機能を含めるものとしている：(1)強化信号伝播、(2)自己解体、(3)新しい素子の生成

各素子は強化信号伝播機能によって、強化信号（報酬値）をより多く受け取るための学習である強化学習を行う。この強化信号伝播機能によってネットワーク上の素子は互いに評価値を算出し合い、互いの重要度を決定できる。さらに、強化信号をより多く受け取れるようにネットワ

ークの構造を変更する機能として、自己解体機能、新しい素子の生成機能を持たせている。これらの機能は、先の強化信号伝播規則によって決定した評価値をもとに実行され、素子単位でのネットワーク構造の生成と淘汰を実現するための機能である。これらの機能によって、ネットワーク構造の自己組織化をオンライン・リアルタイムに行う強化学習の実現ができる。この枠組みによるネットワークの設計は報酬量を安定的に増大させる回路素子の設計に帰着し、その設計がタスク・環境と独立したものとして扱えるため、汎用的な学習システムの構築が期待できることを示している。

第3章では、SONEの教師あり学習への応用と、SONEの基本特性解析試験について述べている。ここでは、軌道学習に関する試験と二重螺旋問題を用いた試験によって、要求仕様のうち特に汎化・抽象化、柔軟性、オンライン性、漸次性の4項目を中心として評価を行っている。また軌道学習に関する試験では、SONEの耐ノイズ特性、追加学習特性、時系列学習特性を解析すると共に、二重螺旋問題を用いた試験によって、汎化能力に関する評価も行っている。これらの試験ではリカレントニューラルネットワーク等を用いた学習法との比較を適宜行っており、SONEを用いた場合の利点についてまとめている。

第4章では、行動創発、柔軟性、オンライン性の両立をシミュレーション上の移動ロボットを用いた衝突回避実験によって確認している。この実験によってSONEが行動創発とオンライン性を両立できることが確かめられている。この実験で用いている移動ロボットは、8個の赤外線センサと2個の車輪を駆動するモータを備えており、フィールド上の壁を感知しながら移動することができる。この実験では、ロボットを制御するプログラムは何も用意せず、ロボットの入出力に対してSONE回路素子を配置する方法をとっている。すなわち、ロボットは全ての必要な回路構造をSONEの自己組織化によって獲得する必要がある。ロボットには直進行動に対する報酬値と壁付近における罰（負の報酬値）を設定し、フィールドを探索させている。この実験の結果、実験開始時にはランダムな行動が多かったロボットの動作が約5分程度で改善されることを示している。ロボットはランダムな行動の経験から、状況に応じた動作を学習・獲得できたものと考えられ、ロボットの行動創発に関する学習がオンラインに行えることを確かめている。

第5章では、SONEの内部には原理的にノイズが発生するという仮説に対し、SONEを構成する各素子に対するノイズ耐性を高めるための手法を提案している。また、ノイズ対策を施したSONEによる移動ロボットの衝突回避実験を行い、その有効性を検証している。SONEの強化学習方式では、教師あり学習の場合とは異なり、強化信号の与え方がネットワーク内部の報酬値に対しノイズを発生させること、従来の自己組織化論理回路へ実

装している素子生成・淘汰の判定方法はこのノイズを軽減させる働きがあることを示している。しかし同時に、ノイズを効果的に除去するためには素子生成・淘汰に関する判定方法を動的に調整しなければならないことも明らかにしている。そこで、その動的な調整方法を対数関数の性質を応用した計算式で各素子に対して実装している。これを実装するにあたって必要となる関数形状の選定には、単純な 3-bit 演算回路を構成する試験を人為的にノイズを与えた環境下で行い、その正答率に従って選定を行っている。ここで得られた耐ノイズ性を向上させた SONE 素子を用いて第 4 章と同様の実験を行い、ロボットの振る舞いが、より精度良く改善できることを確かめている。

第 6 章では本論文の全体に関する考察と結論、さらには今後の展望について述べている。

以上要するに、本論文は、自律型ロボットにおける効果的な制御則を学習によって獲得する手法として、新しい学習制御器 SONE を提案し、その有効性を実験的に示している。SONE が、ネットワーク型の制御回路構造を強化信号伝播規則を導入した学習により獲得する汎用的な仕組みを採用したことなどにより、自律型ロボットの学習制御器に要求される 5 要素である行動創発、汎化・抽象化、柔軟性、オンライン性、漸次性の全てを満たしたことを示している。さらに、これらの有効性を確認した二重螺旋問題評価実験、移動ロボットの衝突回避実験について報告している。

本研究は、自律型ロボットを意識しつつも汎用性のある学習制御器として独自の構造を有する SONE を提案しており、未知環境下で行動するロボットや多様なコミュニケーションが可能なロボットの実現に結び付くだけでなく、さまざまなシステムの知能化に必要な学習機能の効果的な実現方法を提供するものである。このように、ロボット工学、知能機械学、制御工学などの広範な分野に有用となる学習制御器を構築できたことは、ロボットや知能システムの基盤である機械工学、情報工学など関連分野の発展に大きく貢献するものである。よって、本論文は博士（工学）早稲田大学の学位論文として価値あるものと認める。

2008 年 1 月

(主査)	早稲田大学教授	工学博士（早稲田大学）	菅野重樹
	早稲田大学教授	博士（工学）早稲田大学	藤江正克
	早稲田大学教授	工学博士（早稲田大学）	山川 宏
	早稲田大学教授	工学博士（早稲田大学）	高西淳夫