氏　　名	呉本　尭（くれもと　たかし）
授与学位	博　士(工学)
学位記番号	理工博乙第１２９号
学位授与年月日	平成２６年３月４日
学位授与の要件	学位規則第４条２項
研究科，専攻の名称	理工学研究科(博士後期課程)システム設計工学系専攻
学位論文題目	自己組織化ファジィニューラルネットワークを用いた強化学習システムに関する研究
論文審査委員	主　査　山口大学　教授　大林　正直
	山口大学　教授　田中　幹也
	山口大学　教授　浜本　義彦
	山口大学　教授　石川　昌明
	山口大学　教授　松藤　信哉

【学位論文内容の要旨】

人工知能分野に属する機械学習の一つである強化学習に関して、1940 年代に Bellman が提案した動的計画法に基づいたいくつかの学習アルゴリズム、即ち、Actor-critic，Q，Sarsa 等の学習法が提案されている。近年では、これらの学習法を用いた強化学習の応用も盛んに行われてきており、人間を相手にするサービス型ロボット等、従来型の制御では困難な、環境変化に柔軟に対応可能な応用研究も数多くなされてきている。しかしながら、例えば、自然界で発生する現象は時間的・空間的に連続な事象であるのに対し、Q，Sarsa 等の学習法は離散的な現象を対象としている。本論文では、これら、それぞれの学習法に対して、連続的な現象を取り扱えるニューロファジィ型強化学習システムを提案している。そして、これまでは、単独のエージェント（ロボット）を学習の対象とした研究が殆どで有るのに対し、本論文ではエージェント群の効率的な群学習アルゴリズムを提案し、これが単独学習よりも有用である分野を例示している。

本論文の構成は、以下の通りである。

第 1 章では、研究の背景と位置付け、及び論文の構成について述べる。

第 2 章では、知的個体の概念と機械学習の概要を述べ、本論文の研究対象分野であるファジィ推論、ニューラルネットワーク及び強化学習を概説する。特に、観測情報を分類するための自己組織化型ファジィニューラルネットワーク（SOFNN）を提案する。ここで、「自己組織化」とは、入力データ駆動によるネットワークの自動形成を意味する。多次元入力空間による入力に対し、提案手法では、閾値制御及びルール生成規則によって、ファジィメンバーシップ関数や、ファジィルールを生成・結合し、ファジィニューラルネットワーク（FNN）を

構成する。

第3章では、FNNを用いた学習アルゴリズムの異なる三種類の強化学習システムを提案している。

(i) FNNを用いたActor-Critic型強化学習システム（FAC）：

SOFNNの出力に荷重を加え、ネットワークの一層とする状態価値関数Criticと行動価値関数Actorを可塑的に結合する。行動価値関数の出力は、確率探索をもたらす行動選択方策関数に用いられ、知的個体が行動を選択し、出力する。行動の結果による環境の状態の変化と、知的個体に返す報酬や罰を含む時間差分誤差（TD-error）を用いて、SOFNNと状態価値関数・行動価値関数の結合荷重を修正することよって、知的個体が価値の高い状態へ遷移するように、行動方策が修正される。座標情報（離散値及び連続値）を用いた有限マルコフ決定過程(MDP)を持つ目標探索問題のシミュレーションを行い、提案したFACによって構成された知的個体が環境との相互作用の結果より、適切な行動を獲得できることが認められた。

(ii) FNNを用いたQ学習型強化学習システム（FQ）：

システムの構成はFACと異なり、SOFNNの出力は、状態—行動価値関数（Q関数）と可塑的に結合する。行動選択方策関数はQ関数を用いて構成される。1989年にWatkinsにより提案された一般的に利用されているQ学習アルゴリズムと異なり、従来の「TD誤差を直接にQ値の修正に用いる」の代わりに、TD誤差をFACの結合荷重の修正に導入し、システム（FQ）の出力改善につなげる。近傍情報しか得られない部分観測マルコフ決定過程(POMDP)を持つ目標探索問題のシミュレーションを通して、提案したFQによって構成された知的個体が環境との相互作用の結果より、適切な行動を獲得できることが認められた。

(ii) FNNを用いたSarsa型強化学習システム（FS）：

FSの構成はFQと同じであるが、Q学習のTD誤差の計算法と異なり、1994年にRummery & Nirajanにより提案されたSarsa学習のTD誤差式をFSの結合荷重の修正に導入する。また、POMDP下の目標探索問題のシミュレーションを通して、提案したFSによって構成された知的個体が環境との相互作用の結果より、適切な行動を獲得できることが認められた。

また、魚や鳥など生物の群行動のシミュレーションができる行動選択ルールを未知環境における目標探索問題の解法に導入し、知的個体間の適切な距離を保つように、「離れず近すぎず」行動の報酬を個体の価値関数に反映し、「群学習」によって、最適解、または準最適解をより早く発見することを図る。「群学習」の概念をそれぞれの提案強化学習システムに導入した場合と、他の個体との距離をを考慮しない「単独学習」の場合の比較が、シミュレーションの結果によって行われた。

第4章では、不完全観測環境（POMDP）における目標探索問題のシミュレーションを用いて、ランダム探索、従来のQ学習法、従来のSarsa法、及び提案したFAC、FQとFSのそれぞれの学習結果・学習性能の比較を行い、提案法の有効性を確認する。また、提案法におけるパラメータ設定方法について考察する。

第5章では、本論文のまとめと今後の課題について述べる。

【論文審査結果の要旨】

近年、自動車の自動運転への利用等、人工知能研究の重要性が飛躍的に高まってきている。人工知能分野に属する機械学習の一つに強化学習がある。この強化学習は，1940 年代に Bellman が提案した動的計画法に基づいたいくつかの学習アルゴリズム，即ち，Actor-critic，Q，Sarsa 等の学習法が既に提案されている。最近，これらの学習法を用いた強化学習の応用も盛んに行われてきており，人間を相手にするサービス型ロボットの制御等，従来型の制御では困難な，環境変化に柔軟に対応可能な応用研究も数多くなされてきている。しかしながら，例えば，自然界で発生する現象は時間的・空間的に連続な事象であるのに対し，Q，Sarsa 等の学習法は離散的な現象を対象としており，連続事象に対応可能な強化学習方式が求められている．本論文では，これら、Actor-critic，Q，Sarsa，それぞれの学習法に対して，連続的な現象を取り扱えるよう、自己組織化ニューラルネットワークを用いた強化学習システムを提案している．さらに，これまでは、単独のエージェント（ロボット）を学習の対象とした研究が殆どで有るのに対し，本論文ではエージェント群の効率的な群行動学習アルゴリズムを提案し，各エージェントの単独学習よりも有用である分野を例示している。

本論文の構成と内容は以下の通りである．

第 1 章では、研究の背景と位置付け、及び論文の構成について述べている。

第 2 章では、知的個体の概念と機械学習の概要を述べ、本論文の研究対象分野であるファジィ推論、ニューラルネットワーク，そして、本論文で重要な役割を演ずる「自己組織化ファジィニューラルネットワーク (SOFNN)」及び強化学習について記述している。

第 3 章では、既存の強化学習手法 (Actor-Critic, Q, Sarsa) を SOFNN に適合するように発展させ，下記三種類の強化学習システムを段階的・改善的に提案している。

(i) **SOFNN を用いた Actor-critic 型強化学習システム (FAC)**：本システムでは SOFNN の出力に、状態価値関数 Critic と行動価値関数 Actor を可塑的に結合している。行動の結果による環境の状態の変化及び報酬や罰を含む時間差分誤差 (TD-error) を用いて、知的個体が価値の高い状態へ遷移するように、行動方策が修正されている。有限マルコフ決定過程 (MDP) を持つ目標探索問題のシミュレーションを行い、提案した FAC によって構成された知的個体が適切な行動を獲得できていることが認められる。

(ii) **FNN を用いた Q 学習型強化学習システム (FQ)**：本システムの構成は FAC と異なり、SOFNN の出力は、状態—行動価値関数 (Q 関数) と可塑的に結合している。1989 年に Watkins により提案された一般的に利用されている Q 学習アルゴリズムと異なり、TD 誤差を FAC の結合荷重の修正に導入し、システム (FQ) の出力改善につなげている。近傍情報しか得られない部分観測マルコフ決定過程 (POMDP) を持つ目標探索問題のシミュレーションを通して、提案した FQ によって適切な行動を獲得できることを確認している。

(iii) **SOFNN を用いた Sarsa 型強化学習システム (FS)**：FS の構成は FQ と同じであるが、一般の Q 学習の TD 誤差の計算法と異なり、1994 年に Rummery & Nirajan により提案された Sarsa 学習の TD 誤差式を FS の結合荷重の修正に導入している、また、POMDP 下の目標探索問題のシミュレーションを通して、提案した FS によって，適切な行動を獲得できることを示している。

また、魚や鳥など生物の群行動のシミュレーションができる行動選択ルールを未知環境における目標探索問題の解法に導入し、「群学習」によって、最適解、または準最適解をより早く発見することを図っている。さらに，「群学習」を提案の強化学習システムに導入した場合と「単独学習」の場合の比較シミュレーションを行ない、提案法の優越性を示している。

第 4 章では、不完全観測環境 (POMDP) における目標探索問題のシミュレーションを用いて、ランダム探索、従来の Q 学習法、従来の Sarsa 法、及び提案した FAC,FQ と FS のそれぞれの学習結果・学習性能の比較を行い、提案法におけるパラメータ設定方法についての考察、提案法の有効性を確認している。

第 5 章は、本論文のまとめと今後の課題となっている。

公聴会には学内外から多数の参加があり，活発な質疑応答がなされた．その主な内容として，

1．ファジィニューラルネットワークによる状態認知はどのようになされているか，また，状態数の爆発問題は解決されているか.
2．エージェントの最適な行動時系列はどのような形でシステムへ記憶されているか.
3．確率的行動選択に大きな影響を与える温度定数はどのような設定となっているか.

4．学習終了後のシステムにおける行動のロバスト性は担保されているか．

5．学習開始時におけるエージェント間の距離の学習性能への影響について．

等の質問があり，いずれの質問に対しても申請者から的確な回答がなされた．

以上より，本研究は，独創性，信頼性，有効性，実用性ともに優れており，博士（工学）の論文に十分に値するものと判断した．