文 献 紹 介

Lin, L.-J.: Scaling Up Reinforcament Learning for Robot Control, *Proc. of 10th Int. Conf. on Machine Learning,* pp. 182-189 (1993).

強化学習 (reinforcement learning) とは，ある種の学習問題のクラスを指す言葉である．学習者はある環境のなかで行動を起こすエージェント，例えば，自律移動ロボットや動物個体が想定される．学習者は各時間ステップにおいて得られる感覚入力から行動を決定する．実際に取った行動に対して環境から報酬あるいは罰が与えられるが，報酬の大きさは多くの場合，過去数ステップの行動系列に対して決定される．学習の目的は，ある時間長さにわたる報酬の重み和を最大化することである．

本論文は著者 (Long -Ji Lin) の博士論文の要約ともいえるもので，ここ数年間に行った研究のまとめである．基本的な学習機構を解説した後，拡張としての(1)教示(teaching)，(2)階層的な学習，(3)短期記憶の利用について，それぞれの概要が述べられている．

強化学習を実現する戦略として著者は，Qネットと呼ばれる多段ニューラルネットを用いて，Q-learningを実現する方法を用いている．Q-learningとは，感覚入力と行動の組から効用を求めるQ関数を経験から同定し，ある感覚入力に対して効用を最大化する行動を選択するというものである．Qネットでは結合重みの調整を逆伝搬法と時間差分法 (TD(λ)) の組合せによって行う．さらに，経験の再生(experience replay)と著者が呼ぶ方法によって，直接的に報酬を過去の経験に逆伝搬させ，学習の高速化を図っている．

例題として，壁で仕切られた三つの部屋からなる空間を自律的に移動する1台のロボットを考え，計算機シミュレーションによって学習能力の検証を行っている．ロボットの課題は，一つの部屋にある充電装置に自らを接続することである．ロボットは空間全体を見渡すことはできず，地図も持たないため，探検によって適切な経路を見つけ出すことは難しい．

著者は，模範的な移動経路をロボットに示すことによって，学習を容易にする方法を提案している．これは，強化学習における成功例の重要さを示すものでもある．さらに，複数の部屋を通過して目標地点まで到達するために，全体のタスクを(1)壁に沿った移動，(2)出入口の通過，(3)充電装置への接続という三つの要素タスクに分割し，それぞれについてある程度学習した後に，それらの技能の組合せとして全体のタスクを達成する方法を提案している．個々のタスクは上位のものも含め強化学習の枠組みの範囲内にある．

入力から状態が一意に決まらないような環境に対応する方法についても以下の三つを提案している．

(1) 過去数ステップ分の入出力をまとめて入力として扱う．

(2) Qネットにリカレントネットを用いる．

(3) リカレントネットを用いて環境のモデルを構成し，その状態をQネットと入力とする．

移動ロボットのシミュレーション実験では，(2)のリカレントQが最も良い性能を示した．

Q関数の表現には，ニューラルネットのほか，表形式のものや実例に基づく方法などいくつかの手法が他の研究者によって提案されており，入出力データの型あるいは構造に応じたさまざまな関数同定メカニズムを利用することができる．また，経験の再生による高速化は，Michieらの BOXES でとられた方法[Michie 68]や，分類子システムにおける報酬共有(profit sharing)[Grefenstette 88]とほぼ同様であり，特に目新しいものではない．教示については，Clouseらが，系列ではなく1ステップのみを教示するオンライン教示法[Clouse 92]を提案しており，これらを相補的に用いる方法の開発も待たれるところである．階層的な学習は，複雑な課題の実行に不可欠であるが，いまのところ階層化後の学習法については，他にも研究が進められているが，階層自身の発見については，まだ手がつけられていない．

◇ 参 考 文 献 ◇

[Clouse 92] Clouse, J. A. and Utgoff, P. E.: Teaching Method for Reinforcement Learning, *Proc. of 9th Int. Conf. on Machine Learning,* pp. 93-101 (1992).

[Grefenstette 88] Grefenstette, J. J.: Credit Assignment in Rule Discovery Systems Based on Genetic Algorithms, *Machine Learning,* Vol. 3, pp. 225-245 (1988).

[Michie 68] Michie, D. and Chambers, R. A.: Boxes: An Experiment in Adaptive Control, in Dale, E. and Michie, D. (eds.), *Machine Intelligence, 2,* pp. 137-152, Oliver & Boyd, Edinburgh (1968).

〔畝見達夫(創価大学工学部)〕