日本オペレーションズ・リサーチ学会
2004年春季研究発表会

関数近似を用いた強化学習のKheperaロボットへの応用

甲南大学大学院　自然科学研究科　情報・システム科学専攻　*大前　雅裕　OHMAE Masahiro
中山　弘隆　NAKAYAMA Hirotaka

Application of the Reinforcement learning using function approximation to Khepera robot

Abstract : Reinforcement learning is made by the interaction with enviroment. But it is difficult to apply reinforcement learning to robots as it is. Because enormous time of computation is needed. This paper proposes a method which uses function approximation for Reinforcement learning, and employs it to Khepera robot's feature efficiently in an actual environment.

1　はじめに

ロボットの学習方法として強化学習の関心が高まっている．強化学習は環境との相互作用で学習を行うが，実際の環境においてはデータが非常に多くなりロボットにそのまま適用するのは困難である．そこで本稿では強化学習に関数近似を用い，さらにKheperaロボットの特徴を生かしKheperaロボットが実際の環境で学習できる手法を提案する．

2　強化学習

2.1　強化学習とは

強化学習とは，試行錯誤を通じて環境に適応する学習制御の枠組みである．強化学習は報酬というスカラーの情報を手掛かりに学習する．Fig.1で強化学習におけるエージェントと環境の相互作用を示す．

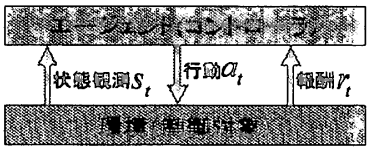

Fig. 1: 強化学習におけるエージェントと環境間の相互作用

強化学習ではFig.1のように，まずエージェントは環境から状態観測S_tを受け取り，行動を決定し，行動a_tを起こす．その行動によって環境は報酬r_tを渡す．これを繰り返し学習する．

2.2　RBFネットワーク

RBFネットワーク (Radial Basis Function:放射基底関数) とは，中心からの距離が離れるにつれて，値が単調に減少（または増加）し，その等高線が超球になる関数のことをいい，代表的なものにガウス関数

$$h(x) = exp(-\frac{\|x-c\|^2}{r^2})$$

がある．本研究ではこのRBFネットワークを用いて関数近似を行う．

2.3　関数近似

強化学習における関数近似の用い方について述べる．強化学習は状態価値関数 ($V(s)$) によって学習する．そこで状態価値関数 ($V(s)$) を

$$V(s) = \sum_{i=1}^{n} w_i h_i(s)$$

によって近似する．nは基底の数，wを重みとする．本当の正しい値を$v(s)$とすると，すべての状態sに対して

$$\frac{1}{2}\{v(s) - V(s)\}^2 \rightarrow Min$$

となるように重みを調節していく必要がある．そこで最急降下法の考え方を用いて，勾配ベクトルの逆の方向dは$\frac{1}{2}\{v(s) - V(s)\}^2$を偏微分したものに$-1$を掛けると，

$$d = \{v(s) - V(s)\}h_i(s) \quad (1)$$

となる.
重みは最急降下法を利用して,

$$w_i \leftarrow w_i + \alpha d$$

を更新していくことによって求めていく. 式 (1) を代入すると,

$$w_i \leftarrow w_i + \alpha\{v(s) - V(s)\}h_i$$

強化学習の考え方により, $v(s) - V(s)$ を TD 誤差で近似すると,

$$w_i \leftarrow w_i + \alpha\{r_{t+1} + \gamma V(s') - V(s)\}h_i \qquad (2)$$

と導ける. この式 (2) を更新することによって最適な重みを求めていく.

3 Khepera ロボットへの応用

3.1 Khepera ロボット

Khepera ロボットには左に 1 つ, 左斜め前に 1 つ, 前に 2 つ, 右斜め前に 1 つ, 右に 1 つ, 後ろに 2 つ, 計 8 つの赤外線近接センサ (左斜め後ろ, 右斜め後ろにはない) が装備されている. このセンサは 0 ～ 1023 の整数値を返し, 障害物が遠くにあるほど小さく, 近くにあるほど大きくなる.

3.2 応用方法

実際の環境で Khepera ロボットの位置を強化学習の状態とすると, Khepera ロボットの正確な位置を判断するのは非常に困難になる. Khepera ロボットの目的は障害物を避ける, 目的地に行く, 物を運ぶなどなので, Khepera ロボットのセンサによって目標物との距離が測れれば十分である. そこで Khepera ロボットのセンサ値を強化学習の状態のパラメータとすると, 1024^8 個の状態を用意することになる. しかし Khepera ロボットに 1024^8 個のデータを持たせることは不可能である. そこで Khepera ロボットのセンサ値を何個かに区切り, センサも 8 個から 4 個に減らすことにより, データ数を減らし学習させてみた. しかしこの方法では近い値のセンサ値なら同じ行動をとったり, 使用しないセンサ部が死角になったりで正しく学習ができなかった. そこで関数近似を用いることでさらに多くのデータ数を扱え, より学習が可能だと考えた. この手法だと RBF ネットワークの基底における重みの数だけデータを用意すれば, あとは計算でそれぞれの状態価値関数 $V(s)$ を求めることができる.

3.3 実験環境

実験環境として, 第 1 の実験は, 迷路の中に Khepera を置きスタートからゴールまでの経路を探索させる. 迷路は Khepera が常に壁を感知できるものとし, ゴールから先には壁を置かずに壁を感知させないつくりにする. Khepera にはセンサしか無いので, センサが壁を感知しなければゴールしたと判断させることができる. どれくらいの学習させればゴールできるようになるか計測する.

第 2 の実験は, 2 台の Khepera を使い追いかけっこをさせる. 逃げる方に本研究を用い, 追う方は学習させずに常にもう 1 台の Khepera を追うようにプログラムする. そして学習用の Khepera は追跡用の Khepera からどれくらい逃げれるかを計測する.

4 結果

実際に Khepera ロボットで実験を行ったところ, センサをすべて使えるため Khepera ロボットの死角が減り, センサの感知する範囲内において, 障害物を回避することができた. また関数近似を使うことによってデータを正確に使えるようになり, 関数近似を使わない方法よりも良く学習できた. 詳細は当日発表する.

5 謝辞

Khepera ロボットについて詳しく教えていただいた, 福井大学工学部知能システム工学科の皆さんに深く感謝の意を表します.

参考文献

[1] Richard S.Sutton and Andrew G.Barto(2001):「強化学習」森北出版

[2] 堀越　俊之 (2002):「適格度トレースと関数近似を用いた強化学習」甲南大学理学部卒業論文

[3] 福井大学リカレント講座 (2002) :「Khepera 入門」