

自律移動エージェントの群行動制御に関する研究

道工大 ○板垣英樹, 道工大 川上敬, 道工大 木下正博, 道工大 大堀隆文

要旨

本報告では自律エージェントが多数存在する群エージェントシステムに対する, 新しい群行動制御に関する理論の構築をめざし, その第一次的な実験として, 群エージェントの穴抜け問題に対して強化学習の適用方法により, 学習性能がどのように異なるのかを実験的に検証し考察する。

1. はじめに

知的な自律エージェントを複数動作させ, 与えられたタスクを実行するマルチエージェントシステムにおいて, エージェント間の協調/競合のような相互作用は理論的に取り扱うには困難が伴うため, 強化学習などの手法を用いて各エージェントの行動を学習により試行錯誤的に組織化する手法が多く提案されている。しかしこのとき, エージェントごとに学習を行い, それぞれのエージェントの行動を学習するため, エージェント数が増加すると, 起こりえる相互作用が非常に複雑となり, なかなか学習が進まないという問題が生じる。

そこで本報告では自律エージェントが多数存在する群エージェントシステムに対する, 新しい群行動制御に関する理論の構築を目指し, その第一次的な実験として, 群エージェントの穴抜け問題に対して強化学習の適用方法により, 学習性能がどのように異なるのかを実験的に検証し考察する。

2. 群エージェントシステム

複数の自律的なエージェントが同時に同じ環境内に存在し, 各エージェントが環境から情報を受け取ると同様に, エージェント間の相互作用により何らかのタスクを達成するシステムをマルチエージェントシステムと呼んでいる。ここではさらに多数の自律エージェントが複雑な相互作用を起こすシステムを群エージェントシステムとして扱い, この群を知的に制御するための方法を議論する。

3. 問題設定

ここでは群エージェントシステムの一つの例として, 多数の自律移動エージェントが一箇所の出口を通過して閉領域から脱出する穴抜け問題を対象問題とする(図1)。図中, 円形のものが自律移動エージェントで, 単位時間に, 上下左右の4方向に1単位長移動することが出来る。ただし, 壁や他のエージェントが接触している場合はその方向には進めないものとする。またこのエージェントは現在位置および, 環境や他のエージェントとの衝突は知覚可能であるが, 出口位置についてはあらかじめ情報は持たないものとする。

このタスクの目的は, 全てのエージェントが出来るだけ早く壁等の障害物を避けながら出口を通過して脱出することである。この問題は自動倉庫等における自律搬送

エージェントのナビゲーションタスクに応用可能で, さらに出口付近でのデッドロックを解消しなければならないという側面をも含んでいる。

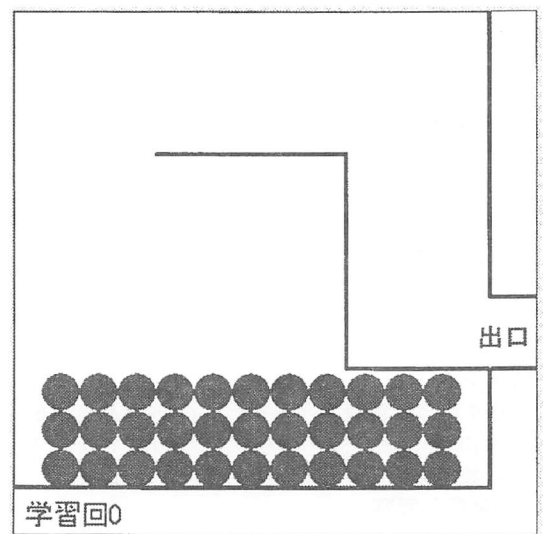


図1 群エージェントの穴抜け問題

4. アプローチ

本報告では上記問題設定に対して, 群エージェントの取るべき適切な行動を導出するために, 代表的な強化学習手法であるQ学習を適用する。

一般的に, エージェントにさせる行動がタスクの遂行にとってどの程度良いかを評価するのは難しく, どのような行動を取るのがよいか各行動に対して教師信号を与えるのは無理がある。そこでエージェントが試行錯誤的に行動し, 環境から与えられる報酬を最大になるように行動を修正してゆくようにする。このような教師信号のない学習手法を強化学習という。

4.1 Q学習の適用

Q学習では時刻 t における状態 s_t においてある行動 a_t をとったときの行動価値関数の値 $Q(s_t, a_t)$ を Q 値と呼び, この Q 値を更新式(1)式に従い学習によって逐次的に更新し, Q 値を最適行動価値関数に近づけようとする手法である。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha [r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t)] \quad (1)$$

ここで、 r_{t+1} はエージェントの行動 a_t の結果、環境から与えられる報酬で、この更新式により、将来にわたって受け取られる報酬が最大になるような最適行動価値関数が獲得できる。また α は学習率、 γ は割引率を表す。

このQ学習を群エージェントの穴抜け問題に適用するにはいくつかの方法が考えられるが、ここでは、状態 s_t をエージェントの座標値とし、行動 a_t は上下左右の4つの方向から1つを選択するものとする。この行動選択はその時点でのQ値から ϵ グリーディ方策により決定する。また、報酬 r_{t+1} は出口に到達した場合のみ+1とし、それ以外は常に0を与えるものとする。

通常、マルチエージェントシステムに強化学習を適用する場合、各エージェントが個別に学習を行い、あるエージェント独自のQ値を獲得する手法が用いられる。しかしながら、ここでは新たな群エージェントの行動制御に関する理論の構築を目指すため、問題空間のフィールド上にエージェント群が共通に利用・更新可能な共有知識を獲得する手法を提案する。すなわち、獲得するQ値はエージェントが個別に持つのではなく、閉領域にフィールドに直接蓄積され、すべてのエージェントが利用するものとする。

5. コンピュータ実験

5.1 実験環境

以上の問題設定に基づき、コンピュータシミュレーションを行う。本実験ではエージェント数を33、エージェントが移動可能な領域は 100×100 とし、エージェントの初期配置は図1の通りとする。Q値の初期値は全て0.0、学習率 $\alpha=0.1$ 、割引率 $\gamma=1.0$ 、 ϵ グリーディ方策に用いる $\epsilon=0.1$ とし、1エピソードにおける最大ステップ数を50000として実験を行った。

5.2 実験結果

コンピュータシミュレーションの結果を以下に示す。図2は学習初期のエージェントの動向で、出口に関する予備知識がないため、各エージェントは他のエージェントや壁と衝突しながら、試行錯誤的にフィールド内を移動・探索している。

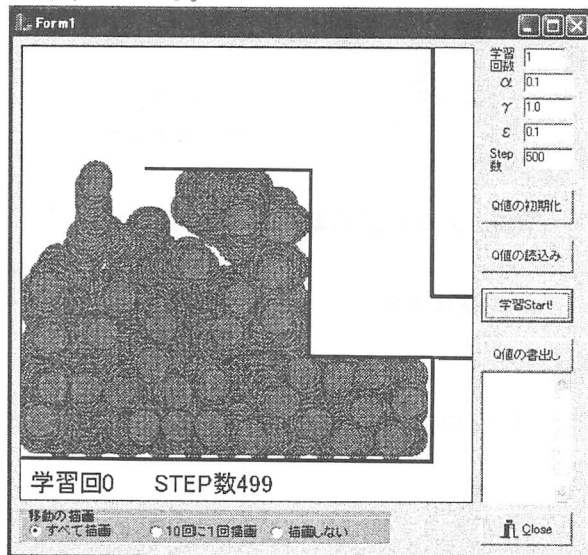


図2 学習初期の群エージェントの挙動

図3は500エピソードの学習を行った後の群エージェントの挙動を表しており、学習によって良好な解が獲得されていることが分かる。また図4、図5はそれぞれ、学習過程における脱出エージェント数と全エージェント脱出までのステップ数を表したグラフである。

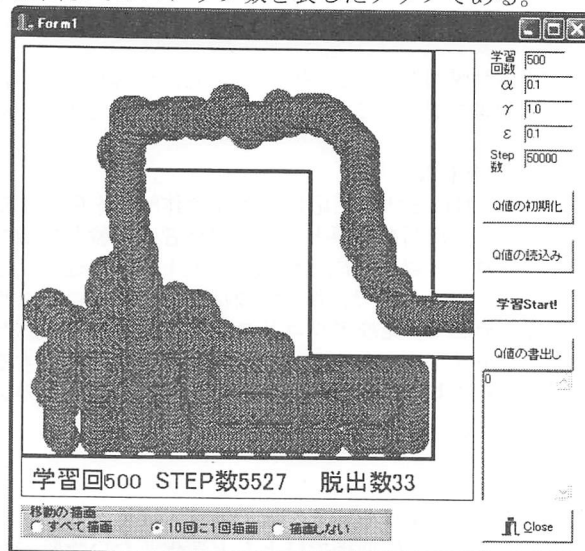


図3 500エピソード学習後の獲得経路

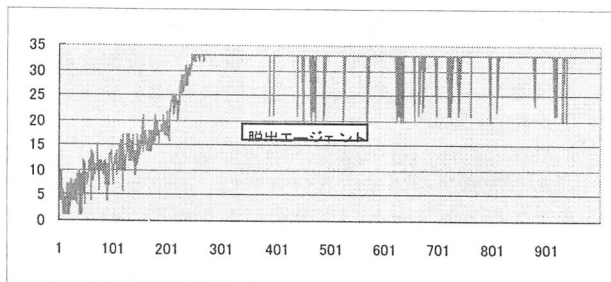


図4 脱出エージェント数

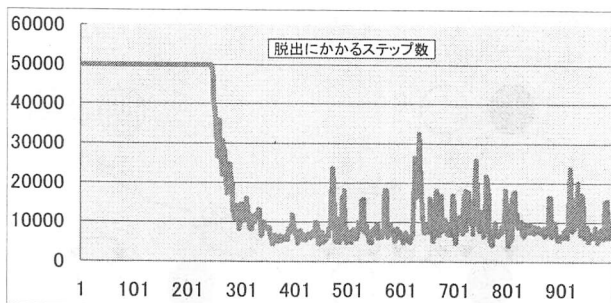


図5 脱出にかかるステップ数

6. おわりに

本報告では群エージェントの穴抜け問題を対象として強化学習を用いた群制御に関する実験を行った。今後は様々な手法との性能比較などを実施する必要がある。

参考文献

- [1]木下正博, 渡辺美知子, 川上敬, 古川正志, 嘉数侑昇: 複数ブロックエージェントの自律行動の獲得に関する研究, 精密工学会誌, vol.68, No.10, pp.1303-08(2002)
- [2]大内東, 山本雅人, 川村秀憲: マルチエージェントシステムの基礎と応用, コロナ社(2003)