

ピアノ問題における強ルールに基づいた協調行動獲得

函館高専 ○吉田 智博、竹原 直美、石若 裕子

要旨

本研究は、強化学習を用いて各エージェントが互いに協調し、自律的に行動系列を学習することを目的とする。ピアノ問題を対象とし、シングルエージェントでは行動系列が複雑になるような問題を、単純な行動をもつエージェント群が協調することで、解決を図る。手法として Q-Learning を用いてシミュレーションを行い、同じ問題をシングルエージェントとマルチエージェントで解き、比較検討した。

1. はじめに

障害物環境下における動作計画問題の一つとして、ピアノ問題がある。ここでは、エージェントが自らの形を持ち、自律的に行動系列を学習させることを目的とする。連続空間におけるピアノ問題の解決策として、マルチエージェントによる協調行動の獲得を試みる。

また、同じ問題についてシングルエージェントでも学習を行い、結果を比較、検討する。

2. ピアノ問題

ピアノ問題とは、荷物を部屋の外に出すために、荷物の回転、および移動など、荷物の動作計画を扱う問題である。この問題は、Schwartz と Sharir^[1]により、荷物が通過する空間の形状が定まっている場合においては、数学的に最適経路の計算が可能であることが証明されている。しかし連続空間において、シングルエージェントでは、回転や移動の方法が複雑になってしまう。そこで、単純な行動を取ることが出来るエージェント群が、マルチエージェントとして協調することによって、複雑な問題を解決する。

3. 問題空間の設定

本研究で用いる空間は、2次元連続空間で、シングルエージェントとマルチエージェントの場合について検討する。シングルエージェントでは、形態を長方形とし、状態として座標をもつ。行動は、長方形の中心を軸として、8方向に回転移動する。Fig.1(a)に、エージェントの移動方向を示す。一方、マルチエージェントでは、長方形の内部に円形のエージェントを2つ持つ。状態として、座標とエージェント自身の角度を持つ。各エージェントは、上下左右4方向に行動する。強ルールにより、実際の行動は、各々の行動の合成で決まり、シングルエージェントよりも複雑な行動を取ることが出来る、Fig.1(b)に、エージェント A0 が右、A1 が上と行動した時の移動位置を示す。

マルチエージェントは、1つの固体内に複数のエージェ

ントが存在する。そのため、エージェント郡の一体が行動を起こした際、固体自体の形は変化しないため、固体全体の挙動が変化する。それを強ルールと呼ぶ。

4. 手法

シングルエージェント、およびマルチエージェントでピアノ問題を解くための手法を以下に説明する。

4.1 ピアノ問題における学習

学習は、Watkins^[2]によって提案された Q-Learning とする。エージェントの更新式は以下のものとなる。

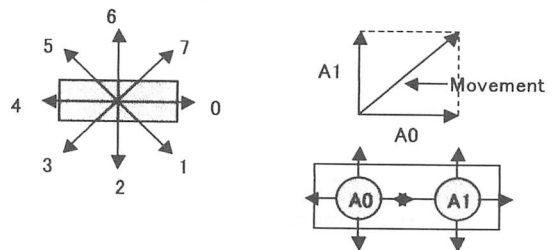
$$Q^{agent_i}(s_t, a_t) = (1 - \alpha)Q^{agent_i}(s_t, a_t) + \alpha \left[r + \gamma \max_{a'} Q^{agent_i}(s_{t+1}, a_{t+1}) \right] \quad \dots (1)$$

ここで、 α は学習率、 γ は割引率、 r は環境からの報酬、関数 \max は行動集合に対して、最も Q 値の高い行動を選択するものである。

時刻 t における状態 a_t の選択方法は以下のものとする。

$$a_t = \arg \max_i Q^{agent_i}(s_t, a_t) \quad \dots (2)$$

ここで、関数 \max は、全エージェントのすべての行動の中から、最も評価価値の高いものを選ぶ。



(a) Single Agent

(b) Multi Agents

Fig.1 The shapes and The direction of Agent's action

4.2 RBF

Q 値の更新は、空間が連続であるという点から、ラジアル基底関数 (RBF^[3]) を用いる。

座標(x,y)に対する RBF の値は以下の式で定義される。

$$gaussian(x, y) = \exp\left\{-\frac{1}{2}\left(\frac{(cx-x)^2}{\sigma_x^2} + \frac{(cy-y)^2}{\sigma_y^2}\right)\right\} \dots (3)$$

cx,cy は中心座標で、長方形の中心座標とする。

σ_x^2 および σ_y^2 は、それぞれ長方形の長辺と短辺に対する分散である。更新範囲は、長方形内とする。

マルチエージェントについては、以下の式で定義される。

$$gaussian_i(x, y) = \exp\left\{-\frac{1}{2}\left(\frac{(cx_i-x)^2 + (cy_i-y)^2}{\sigma^2}\right)\right\} \dots (4)$$

cx_i, cy_i は、各エージェントの中心座標である。 σ^2 は、分散である。更新範囲は各エージェントの円内とする。

5. 実験

Q-Learning を用いた協調行動獲得について検討するため、以下の設定した問題空間においてシミュレーションを行った。

1. 通路が斜めに配置されていて、通路に対して水平方向からしか入れない空間 (200×200)

2. 中心を軸としての回転が難しい空間 (115×80)

実験に用いたパラメータは、以下のように設定した。

エージェントの形状を長方形 (縦 20×横 40) とする。マルチエージェントは、半径 8 の円形とし、中心からの距離 10 の位置に、左右 2 つ配置した。Q-Learning のパラメータは、学習率 $\alpha = 0.1$ 、割引率 $\gamma = 0.9$ 、報酬 $r = 1$ (ゴール到達時)、方策は ϵ -greedy ($\epsilon = 0.05$) とする。エージェントの移動量 $\Delta t = 4$ 、RBF の分散 $\sigma_x = 4$ 、 $\sigma_y = 2$ 、 $\sigma = 4$ 、最大ステップ数 2000、最大エピソード数 1000 とした。以上の条件により実験を行い、各エージェントの軌跡を描いたものを Fig.2 および Fig.3 に示す。

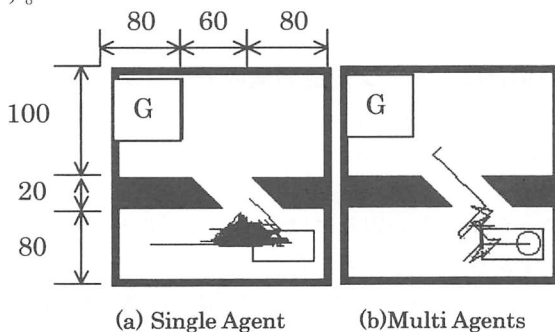


Fig.2 Environment setting 1 and trajectory of rectangle agent. (After 500 episodes)

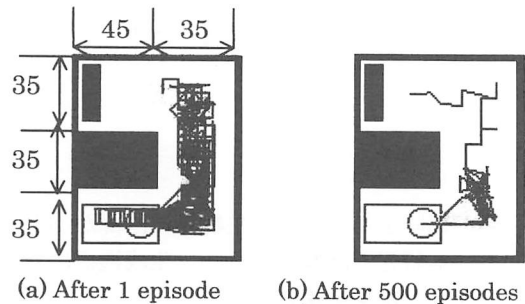


Fig.3 Environment setting 2 and trajectory of rectangle agent. (After 500 episodes)

Fig.2 および Fig.3 に表示されている長方形は、各エピソードに対する初期位置である。Fig.2 に、(a)500 エピソード終了後のシングルエージェントの移動軌跡、(b) 500 エピソード終了後のシングルエージェントの移動軌跡を描いた。Fig.2(a)より、シングルエージェントはゴールにたどり着くことが出来なかった。理由として、シングルエージェントは回転軸を中心にしか取ることが出来ないため、Fig.2 のような問題空間においては、一方向の行動系列を取らない限り通路を通り抜けることは出来ない。それに対しマルチエージェントでは、Fig.2(b)が示すように、一度進んでから回転して、通路に侵入することが出来る。これにより、8 方向移動可能なシングルエージェントよりも、4 方向移動可能なマルチエージェントのほうが、問題空間が複雑になった際に、解決が有効であるといえる。Fig.3 の結果より、(a)は、エージェント間に差がないが、エピソードが増えるにつれて、同じエージェントがゴールに着くようになった。これは、一定のエージェントをゴールに向かうように相手のエージェントが行動するようなエージェント間に協調行動が創発されたといえる。

6. おわりに

本研究では、Q-Learning を用いたマルチエージェントによるピアノ問題の協調行動獲得の手法を提案し、シミュレーションにより、本手法が、ピアノ問題の解決に対して有効であることを示した。今後は、現在は長方形であるエージェントの形状を、より複雑な形状とする。その協調行動獲得のために、エージェントの配置や、状態の持ちかたについて検討することが必要である。

参考文献

- [1] Jacob T. Schwartz, Micha Sharir: On the Piano Movers' Problem: II. General Techniques for Computing Topological Properties of Real Algebraic Manifolds, Planning, Geometry, and Complexity of Robot Motion, pp. 51-96,
- [2] Watkins, C.J.H., and Dayan, P.: Technical note: Q-learning, Machine Learning, Vol.8, pp.55-68, 1992
- [3] 丸山稔, Radial Basis Functions を用いた学習ネットワークニューロコンピューティングに対する新しいアプローチ, システム/制御/情報, Vol.36, No.5, pp. 322~329, 1992