2410 ゲームの理論を用いたマルチエージェントシステムの構築

Constructing Multi-Agent Systems using Game Theory

○中田 篤 (姫路工業大) 小西 康夫 (兵庫県立大)
石垣 博行 (兵庫県立大)

Atsushi Nakata, Himeji Institute of Tech. Yasuo Konishi, University of Hyogo
Hiroyuki Ishigaki, University of Hyogo

In this paper proposes to construct the multi agent system with the game theory, and to control the crowd robot cooperatively. The effectiveness of the proposed method was verified by the pursuit evasion simulation which uses plural robots. In this simulation, pursuit robots are using the differential game to pursue the escape robot. And, pursuit robots are using non-zero-sum two-person game to select the pursuit action and the escape obstruction action. As a result, the pursuit robot acted as a crowd in consideration of the action of other pursuit robots. Moreover, when you compare the usual method with the proposed method, it was clarified that the proposed method was effective.

key words : game theory, multi agent system, crowd robot, cooperated operation

1. はじめに

ロボティクスは様々な分野の技術が集約されており、コンピュータ技術の発達と相まってその進展はめざましい。またロボティクスの発達とともにロボットの使用される環境も、これまでの作業とロボットの関係が1対1の関係ではなくなり、複雑化・多様化を見せている。そのような環境下において、単一ロボットシステムではタスクの達成が困難となるため、複数のロボットを群として制御すること(以下、マルチエージェントシステム)で協調動作を実現する研究が行われている。マルチエージェントシステムには、

- ・環境の変化に対する適応性
- ・様々な作業要求に対する柔軟性
- ・一部の故障が全体の故障に至らない耐故障性
- ・エージェントが並列的に動くことによる作業効率の向上

などのメリットがある⁽¹⁾。従来では、このマルチエージェントシステムの協調動作獲得のために強化学習が用いられてきた^(2,3)。しかし、強化学習には学習に時間がかかるなど問題があり、ロボット(以下、エージェント)の数が増えたとさらに学習時間が増加するため、この問題は致命的なものとなっている⁽⁴⁾。

そこで本研究では、強化学習にかわる制御手法としてゲームの理論を用いたマルチエージェントシステムを構築し協調動作の実現をする手法を提案する。提案法の有効性は協調動作が必要となる複数のエージェントによる捕捉問題を取り上げ、追跡・逃走シミュレーションを行うことで検証した。このシミュレーションにおいて、エージェントの追跡に微分ゲームを適用している。また、エージェントの追跡行動と逃

走妨害行動の選択に非ゼロ和2人ゲームを適用している。

従来の強化学習との比較も、追跡逃走シミュレーションを行い、提案法が従来法の問題を解決できているか検証した。

2. ゲームの理論について

ゲームの理論とは1928年にフォン・ノイマンによって創始された、複数の行動主体が利害の衝突する状況において相互の依存関係を定式化し、その構造を分析することでとるべき行動を決定する理論である⁽⁵⁾。

本研究では、マルチエージェントシステムの行動の最適化にゲームの理論を適用することで協調制御を実現する。協調動作が必要なマルチエージェントシステムの一例として、複数の追跡エージェントによる逃走エージェントの捕捉問題を取り上げる。捕捉問題とは、複数の追跡エージェントが逃走エージェントを追跡・捕捉するもので、追跡エージェントは逃走エージェントを効率よく捕捉するために他の追跡エージェントの行動を考慮しながら自分自身の行動を最適化する必要がある。

3. 追跡・逃走シミュレーションとそのモデル

本研究では、2輪車両型移動ロボットをシミュレーションモデルとして、追跡エージェント(Pursuer)3台が逃走エージェント(Evader)1台を追跡し、捕捉する追跡・逃走シミュレーションを行う。図1にシミュレーションモデルを示す。

2輪車両型移動ロボットのシステム方程式は、式(1)で表される。ここで、状態変数 $\mathbf{x} = [x, y, \theta]^T$ はモデルの重心位置である $x[m]$, $y[m]$ 及び姿勢角 $\theta [rad]$ であり、制御変数 $\mathbf{u} = [V, \omega]^T$ はモデルの移動速度 $V [m/s]$, 及び角速度 $\omega [rad/s]$ である。本シミュレーションでは追跡エージェントと逃走エ

Fig.1 Simulation model

$$f(x, u, t) = \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \end{bmatrix} = \begin{bmatrix} \cos \theta & 0 \\ \sin \theta & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} V \\ \omega \end{bmatrix} \quad (1)$$

エージェントはお互いの状態を観測しているものとする。また、逃走エージェントの捕捉条件は、逃走エージェントと追跡エージェントとの間の距離が0.25 [m]以下になったときとしている。

4. 微分ゲームによる最適制御

追跡・逃走シミュレーションにおいて、追跡エージェントと逃走エージェントの間の距離が重要となる。追跡エージェントはこの距離の最小化を、逃走エージェントは最大化をはかりながら連続的に行動しなければならない。このエージェントの行動を連続的に最適化するために微分ゲームを用いる。

微分ゲームとは、ゲームの開始と終了が明確で段階的に行われる通常のゲームと異なり、ゲームの間の時間間隔を極限にまで小さくすることでゲームを連続的に捉えるものである^(6,6)。本追跡・逃走シミュレーションにおいて、追跡エージェントと逃走エージェントの間の距離を評価関数として、追跡エージェントは評価関数の最大化を、逃走エージェントは評価関数の最小化をはかるものとする。式(2)～(4)に行動最適化のための評価関数を示す。

$$J = S[x(\Delta t), \Delta t] + \int_0^{\Delta t} L[x, u_p, u_e, t] dt \quad (2)$$

$$S[x(\Delta t), \Delta t] = \frac{1}{2} \{ (x_p - x_e)^2 + (y_p - y_e)^2 + (\theta_p - \alpha)^2 + (\theta_e - \alpha)^2 \} \quad (3)$$

$$L[x, u_p, u_e, t] = \frac{1}{2} (V_p^2 + \omega_p^2 + V_e^2 + \omega_e^2) \quad (4)$$

追跡エージェントはこの評価関数の値を最大化する解 $u_p = [V_p, \omega_p]^T$ 、逃走エージェントは最小化する $u_e = [V_e, \omega_e]^T$ を求めることでそれぞれの最適行動が得られる^(6,7)。

この各エージェントに対する最適解 u_p, u_e を得るために、ハミルトニアン $\lambda(t)$ を導入した関数 H に変分法を適用する。

$$H(x, u_p, u_e, \lambda, t) = L(x, u_p, u_e, t) + \lambda^T \cdot f(x, u_p, u_e, t) \quad (5)$$

式(5)を解くことで、時刻 t [s] における最適解 $u_p = [V_p, \omega_p]^T$ 、 $u_e = [V_e, \omega_e]^T$ は次のように求められる。

$$V_p(t) = -\frac{1}{2} \{ (x_p(t) - x_e(t)) \cos \theta_p + (y_p(t) - y_e(t)) \sin \theta_p \} \quad (6)$$

$$\omega_p(t) = -\frac{1}{2} (\theta_p(t) - \alpha(t)) \quad (7)$$

$$V_e(t) = -\frac{1}{2} \{ (x_p(t) - x_e(t)) \cos \theta_e + (y_p(t) - y_e(t)) \sin \theta_e \} \quad (8)$$

$$\omega_e(t) = \frac{1}{2} (\theta_e(t) - \alpha(t)) \quad (9)$$

5. 協調動作の実現

全ての追跡エージェントが前章の微分ゲームで求めた最適解 u_p, u_e のみで行動すると、各追跡エージェントが逃走エージェントを個々の判断だけで追跡することになる。これは、1対1の捕捉問題とかわらないため、逃走エージェントを効率よく捕捉することができない⁽⁸⁾。よって、追跡エージェントは追跡行動を最適化するだけでなく、逃走エージェントの逃走を妨害する行動を取ることで群れとしての行動を最適化する必要がある。この追跡行動と逃走妨害行動の2つを状況に応じて選択するための手法として、本研究では非ゼロ和2人ゲームを適用する。

今、シミュレーション中のある時刻において、追跡エージェント A と B の間に、表1のような報酬行列が存在したとする。各追跡エージェントは追跡行動(添え字: 1)か逃走妨害行動(添え字: 2)のどちらの行動を選択すべきか、マクシミン戦略(maximin strategy)に従い行動を選択する。

Table.1 Reward matrix

Reward for agent A			Reward for agent B		
	b ₁	b ₂		b ₁	b ₂
a ₁	3	1	a ₁	0	3
a ₂	-1	2	a ₂	2	1

マクシミン戦略とは、追跡エージェント A の場合、A の報酬行列において追跡行動 a₁ をとったときの最小の報酬である1と、逃走妨害行動 a₂ をとったときの最小の報酬である-1を比較し、大きな報酬が得られる追跡行動 a₁ を選択するものである。このとき、追跡行動 a₁ を選んだことで、追跡エージェント B が追跡行動 b₁、逃走妨害行動 b₂ のどちらを選んだとしても、追跡エージェント A は最低でも1の報酬を得ることができる。追跡エージェント B も同様に B の報酬行列においてマクシミン戦略に従い、各行動の最小値を比較しその中で最大の報酬が得られる逃走妨害行動 b₂ を選択することで、最低でも1の報酬を得ようとする。追跡エージェントがそれぞれマクシミン戦略によって行動を選択した結果、追跡エージェント A は追跡行動を、追跡エージェント B は逃走妨害行動を取ることで群れとしての行動を最適化する。

報酬行列の中の各報酬は、式(10)に示すように全追跡エージェントの重心位置と逃走エージェントまでの距離としている。

$$G_x = (x_{p1}(t) + x_{p2}(t) + x_{p3}(t) + \dots + x_{pn}(t)) / n$$

$$G_y = (y_{p1}(t) + y_{p2}(t) + y_{p3}(t) + \dots + y_{pn}(t)) / n$$

$$r_{ij} = -\{(x_e(t) - G_x)^2 + (y_e(t) - G_y)^2\} \quad (10)$$

この距離が小さければ追跡エージェントは逃走エージェントを取り囲んでいることとなり、効率の良い捕捉が可能となるためである。

6. マルチエージェントシステムによる協調制御

6.1 マルチエージェントシステムの構築

本研究では、微分ゲームによる移動の最適化と、非ゼロ和2人ゲームによる協調動作を統合することでゲームの理論を用いたマルチエージェントシステムを構築し協調制御を行う手法を提案する。マルチエージェントシステムを構築するときに注意すべき点は、エージェントの数が增多ることによってシステムが複雑にならないようにすることである。システムの複雑化は作業効率の低下や通信の停滞による信頼性の低下を引き起こす。従って、できる限り少ない制御ルールによってマルチエージェントシステムを構築するためにBoidモデルを参考にしている。Boidモデルとは複数のエージェントに数個の単純なルールをあてはめることで、群れの動きを表現する手法である⁹⁾。本構築システムでは、以下の3つのルールを追跡エージェントにあてはめることで追跡エージェントを1つの集団として制御するようにした。図2にその概念図を示す。

1. 整合(Coordination)

追跡エージェントは逃走エージェントの方向を向くように行動する。

2. 引き離し(Separation)

追跡エージェントは他の追跡エージェントに近すぎないように、最も近くにいるエージェントの反対方向へ移動する。

3. 結合(Cohesion)

追跡エージェントは、追跡エージェント全体の重心位置に逃走エージェントが存在するように行動する。

整合(Coordination)は微分ゲームによる移動の最適化によって行われており、引き離し(Separation)は、逃走エージェントと各追跡エージェントの間の距離に応じてルールを適用している。また、結合(Cohesion)は、式(10)に示す評価関数を定め、この評価関数の値を表.1における報酬*r*として非ゼロ和2人ゲームによって行動を決定する。

Fig.2 Boid model

6.2 エージェントの行動ルール

追跡エージェントの行動ルール

追跡エージェントは非ゼロ和2人ゲームに従い、状況に応じて追跡行動と逃走妨害行動のどちらの行動を取るか決定している。ここで逃走妨害行動について説明する。逃走エージェントの逃走を妨害するために、追跡エージェントは逃走エージェントの進行方向に先回りする。このとき、先回りの位置を図3に示すように決定している。

Fig.3 Predict algorithm

追跡エージェントは、逃走エージェントの1ステップ前と現在の位置から逃走エージェントが直進したと仮定する。その直進方向と自分の進行方向が交わる点を予測地点として移動することで、逃走エージェントの進行方向をさえぎる。この結果、逃走エージェントは逃走方向に追跡エージェントがいることとなり、逃走が妨害される。

逃走エージェントの行動ルール

逃走エージェントは複数の追跡エージェントから逃げるために、1台1台の追跡エージェントに対して微分ゲームによって逃走方向を決定し、距離に応じて重みをつけた逃走ベクトル(V_p)を作る。それを合成することで最終的な逃走方向を決定する。図3にその概念図を示す。重みは式(11)によって決定している。この重みは、捕捉条件である逃走エージェントと追跡エージェントとの間の距離が0.25[m]以内で重みを最高の1とし、追跡エージェントと逃走エージェントの間の距離が大きくなるほど重みは小さくなっていく。

$$W_p = \begin{cases} 1 & 0 \leq L_p \leq 0.25 \\ e^{-0.7L_p + 0.175} & 0.25 \leq L_p \end{cases} \quad (11)$$

Fig.4 Evasion movement

7. シミュレーション結果

図5に、微分ゲームと非ゼロ和2人ゲームを統合した、案法であるゲームの理論を用いたマルチエージェントシステムによる追跡・逃走シミュレーションの結果を、図6に微分ゲームによる追跡行動のみを行う追跡エージェント群のシミュレーション結果を示す。それぞれ、(a)はシミュレーション中の各エージェントの軌道を、(b)はシミュレーション中の各追跡エージェントと逃走エージェントとの間の距離の変化を示している。

Fig.5 Using game theory

Fig.6 Pursuit behavior only

図6を見ると、追跡行動のみの場合、追跡エージェントは他の追跡エージェントの行動を考慮していないため、逃走エージェントを各追跡エージェントが並んで追いかけて、捕捉までに21ステップかかっている。この結果追跡行動のみの場合、効率よく捕捉することができないことがわかった。一方、提案法の結果である図5を見ると、追跡エージェントが逃走エージェントを取り囲むように移動しており、捕捉までに12ステップしかかかっていない。つまり逃走エージェントを効率よく捕捉できていることがわかる。これより提案法によって協調動作を実現し逃走エージェントを効率よく捕捉できていることがわかった。

8. 強化学習との比較

提案法であるゲームの理論を用いたマルチエージェントシステムによる協調制御法と、従来研究されてきた強化学習によるマルチエージェントシステムと比較を行う。比較の対象は強化学習のQ-learningによって協調動作を学習したマルチエージェントシステムである。

8.1 Q-learning による協調動作の獲得

Q-learning とは、各状態 $S \in \{s_1, s_2, s_3, \dots\}$ において選択可能な行動 $A \in \{a_1, a_2, a_3, \dots\}$ の中で最も行動価値関数の高い行動 $\max_a Q(s, a)$ を選択するように学習を行う手法である。式(12)によって $Q(s, a)$ の要素の値を更新していく。

$$Q(s(t), a(t)) \leftarrow (1 - \alpha)Q(s(t), a(t)) + \alpha[r(t) + \gamma[\max_a Q(s(t+1), a(t))]] \quad (12)$$

本シミュレーションにおいて図7に示すように、状態 S はエージェントの周りを12等分し、他のエージェントがどの位置にいるか確認している。また行動 A は $\pm\pi/4$ の範囲を7等分し、選択した方向へ移動するようにしている。

Fig.7 Figure of states and actions

報酬は提案法と同様、全追跡エージェントの重心位置と、逃走エージェントの間の距離としている。これより、他の追跡エージェントと逃走エージェントの位置を確認し、その状態において行動 A の中から逃走エージェントに最も近づき、大きな報酬を得るための行動を学習していく。

8.2 比較シミュレーション

前章と同様、追跡エージェント3台と逃走エージェント1台による追跡・逃走シミュレーションを行い従来法と提案法の比較を行った。図8は横軸に学習回数、縦軸に追跡エージェントのシミュレーション中における1ステップあたりの行動の価値をとったグラフである。

Fig.8 Relation between learning time and action value

この図から、学習回数が6000回を越えると行動価値の値がほぼ一定値に収束している事がわかる。これは、学習回数が6000回を越えると、価値の高い行動を選択するようになり、1ステップあたりの行動価値が同じになってきていることを示している。また、Q-learningによって10000回強化学習

を行った追跡・逃走シミュレーションの結果を図9に、提案法であるゲームの理論を用いたマルチエージェントシステムによるシミュレーションの結果を図10に示す。前章同様(a)はシミュレーション中の各エージェントの軌道を、(b)はシミュレーション中の各追跡エージェントと逃走エージェントとの間の距離の変化を示している。

Fig.9 using reinforcement learning

Fig.10 Using game theory

シミュレーション結果 (a) を比較すると、10000 回強化学習を行ったエージェントの軌道は提案法のエージェントの軌道と酷似していることがわかる。また、逃走エージェントの捕獲にかかるステップ数もシミュレーション結果 (b) から従来法が提案法とほぼ同じステップ数で細くしていることがわかる。これより、強化学習は 6000 回以上学習を行うことで、提案法と同じ行動を取るようになることがわかった。よって、提案法は、学習なしに協調動作を実現していることから、提案法によって従来の強化学習の欠点である学習時間の問題を解決できたと言える。

7. おわりに

本研究では、ゲームの理論を用いたマルチエージェントシステムを構築することで、協調動作を実現することを提案した。この結果、非ゼロ和 2 人ゲームを適用することで追跡エージェントは群れとして行動した。よって提案法が有効であることを示せた。また提案法と、従来法である強化学習を追跡・逃走シミュレーションによって比較を行った。その結果、提案法が学習時間なしに協調動作を獲得できていることから、強化学習の問題点である学習時間の問題を解決できたことがわかった。

参考文献

- 1) 佐藤 重彦, 野澤 昭雄, 井出 英人: 感情モデルを適用した自律移動型ロボット群の行動特性, 電気論文学会誌 C, Vol.124, No.7, 2004
- 2) 内部 英治, 浅田 稔, 細田 耕: 複数の学習するロボットの存在する環境における協調動作獲得のための状態空間の構成, 日本ロボット学会誌, Vol.20, No.3, 2002
- 3) 大友 照彦, 小平 実: ニューラルネットワークを用いた移動ロボットの動的環境における走行制御, システム情報学会論文誌, Vol.14, No.2, 2001
- 4) 新井 義和, 藤井 輝夫, 浅間 一: 群ロボット環境における局所的通信に基づく衝突回避, 日本ロボット学会誌, Vol.19, No.1, 2001
- 5) 西田 俊夫: ゲームの理論, 日科技連, 1973
- 6) Puskin Kachroo: Optimal Control for a Two Player Dynamic Pursuit Evasion Game; The Herding Problem, Virginia Polytechnic Institute and State University, 2002
- 7) 大塚 敏之, 石谷 雅宏: 非線形四輪車両モデルの Receding-Horizon 微分ゲーム, 日本機械学会論文集 C 編, Vol.66, No.652, 2000
- 8) 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, 2002
- 9) 小堀研一, 柏原巧: 3 次元空間における群集アニメーション, 画像ラボ, 2004
- 10) 刈谷 学, 鎌野 琢也: 捕獲問題における群ロボットの協調動作の実現, 電気論文学会誌 C, Vol.124, No.2, 2004