自律性と調和性のスイッチングメカニズムによる群れ行動獲得に関する研究

北海道科学大学　○平井　鷹行，北海道科学大学　木下　正博，北海道科学大学　川上　敬，
北海道科学大学　西川　孝二，株式会社トラストフォース　柴田　将利

要 旨

自律的なエージェント群が群れで行動するとき，個々のエージェントは目的を達成するための行動パターンを構築する．このとき貪欲な振舞いを主とする行動方策の他に，調和的な振舞いを両立した方策を適用し，ある種の行動パターンを構築することで，デッドロック現象や過学習などの問題に対応できると考えられる．本研究では，自律性と調和性を両立した行動獲得のための方法論を提案する．

1.　はじめに

自律的なエージェント群の行動獲得において社会生物学における知見は極めて重要なヒントを与えていると考えられる．例としてシェリングが提示した人の住み分け問題を取り上げたシェリングのゲームでは，人種の違いに寛容的であってもマイノリティに属する恐れから自己組織的に分かれてしまうことを証明した．この結果から，個々のエージェントが持つ自律性と調和性を両立したメカニズムを実装し，自己組織的に行動パターンの構築を行う方法が考えられる．本研究では，群れ行動獲得のためにスイッチングメカニズムを提案する．

2.　自律性と調和性

ある系に複数の自律性を持ったエージェントが存在し，相互に作用するとき，環境は極めて複雑な様相を呈する．このことから，最適な知的制御システムの実装が必要となる．その一つにグリーディな方策が取られる．この方策は，最適な値を貪欲に選択しつづけることで最適な振舞いへと収束していくが，複数のエージェントが存在する系においては，常に最適な方策であるとは考え難い．即ち，複数のエージェントが存在する系では，グリーディ方策を基本的な行動戦略とし，状態に合わせて調和的な振舞い，即ち，衝突を避ける行動を取るような振舞いの両立し，振舞いを収束させるような方策が考えられる．本研究では，自律性と調和性を両立した行動を獲得するために，スイッチングメカニズムを実装し，議論する．

3.　スイッチングメカニズム

個々のエージェントが，目的の達成に応じて何らかの報酬が与えられるとき，エージェントは得られる報酬を最大化するために行動を収束していく．しかし，群れの中で情報を共有するシステムを備えていない場合，デッドロック現象や過学習などの問題が発生しやすくなると考えられる．このことから，エージェントが目的を達成するために行動する中で，何らかの状態に遭遇したときに方策を切り替えるようなスイッチングメカニズムが考えられる．本研究では，エージェントとの衝突時には異なる方策を適用するよう設定し，実験を行う．

4.　シミュレーション概要

この節では，本研究で行ったシミュレーションの概要について述べる．

4.1　問題設定

実験環境を図 1 にて示す．実験環境は 250×250 の正方形で構成され，さらに 10×10 のタイルが敷かれている．このタイルを状態とする．中央には障害物が設けられ，上下の部屋には目的となるゴールが設定されている．エージェント群は，最初に設定されているゴールに到達した際，次のゴールに到達するように設定される．この行動を繰返す．

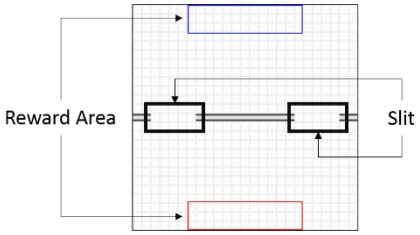

Fig1 Simlation Environment

4.2　エージェント

エージェントは 8 方向に移動することができ，エピソード開始時には上下の部屋に初期配置される．また，エージェントは目的を達成すると色を変更する．その図を図 2 に示す．また，エージェントの移動量はタイルのサイズよりも小さい移動量を取る．その図を図 3 にて示す．この移動量とすることで，エージェントは滑らかに移動することが出来る．

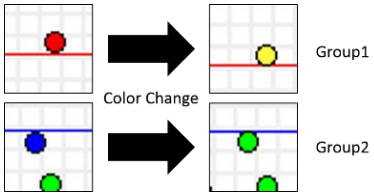

Fig2 Example of agent color change

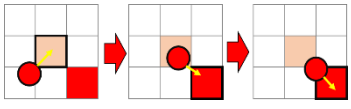

Fig3 Example of agent move

4.3　強化学習

個々のエージェントには強化学習の一つである Q 学習を実装する．Q 学習は以下の状態行動価値関数の更新式(1)を更新していくことで学習を進める．

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

基本的な行動方策は ε -グリーディ方策を適用する．その式は(2)の通りである．

$$Q^*(s, a) = \max_\pi Q^\pi(s, a) \quad (\forall_s \in S, \forall_a \in A) \quad (2)$$

エージェントとの衝突時ではあるステップ間，また ε 以下であったときにはランダム方策を適用する．ランダム方策は次の式(3)の通りである．

$$\pi(s, a) = 1/8 \ (\forall_s \in S, \forall_a \in A) \quad (3)$$

これらの式を合わせて衝突判定$collide$を設定する．その式は(4)の通りである．

$$collide = \begin{cases} True: & 1/8 \ (\forall_s \in S, \forall_a \in A) \\ False: & \max_\pi Q^\pi(s, a) \quad (\forall_s \in S, \forall_a \in A) \end{cases} \quad (4)$$

衝突判定の結果$False$であれば ε -グリーディ方策を，$True$であればランダム方策を適用する．

4.4　定量化

あるエージェントがタイルを通過した際の回数を正規化して定量化を行う．正規化は以下の式(5)の通りである．

$$N = X - x_{min} / x_{max} - x_{min} \quad (5)$$

これにより，グラフ間での比較を容易に行うことが出来る．

5.　シミュレーション

実際に行ったシミュレーションの結果を示す．エピソードは 100，上限ステップ数は 10,0000，報酬は目的達成時に 100 を与え，壁に衝突した際には-10.0 の罰則を与える．学習率は 0.01 で割引率は 0.9 で行う．エージェントは 10 台でグループを形成し，総数 20 台で行う．また，エージェント衝突時のステップは 20 ステップで設定する．図 4 はエピソード 50 時の軌跡とエピソード 100 時の軌跡を示し，図 5 は最終エピソードにおける結果を示す．

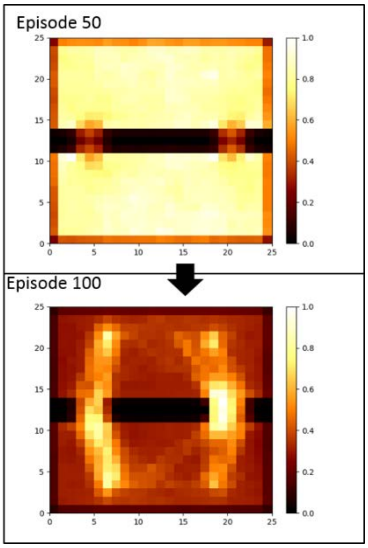

Fig4 Agent of action trajectory

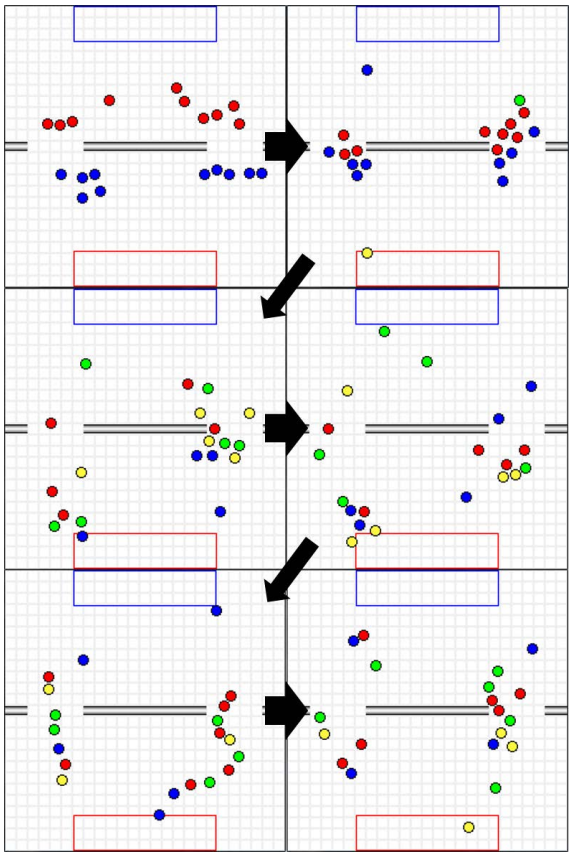

Fig5 Result simulation

6.　考察

本研究では，Q 学習を用いて二つの方策を設定し，衝突時に切り替えるスイッチングメカニズムの実験を試みた．図 4 に示されている通り，エピソード 50 時ではランダムに動いていたが，最終エピソード時にはスリットの間をすり抜けるように行動を獲得できていた．また，実際のエージェントの行動結果を示す図 5 では，スリット空間で衝突し続けていたが，衝突時にランダム方策へと 20 ステップ間適用しつづけることにより，ある種の循環行動が形成されたと考えられる．

7.　おわりに

個々のエージェントが群れを形成し行動する環境下において，群れが報酬の最大化を行うためには，目的を達成するための最適な行動と最適ではない行動の配分を適切に行い，連続的な行動パターンの構築が必要であると考えられる．本研究では，基本的な行動方策と衝突時の行動方策を設定したスイッチングメカニズムを提案し，実験を通してその可能性を示した．

参考文献

[1] マーク・ブキャナン，“人は原子，世界は物理法則で動く　社会物理学で読み解く人間行動”,白揚社,2009 年

[2] MasahiroKINOSHITA,MichikoWATANABE,TakashiKAWAKAMI,HiroshiYOKOI,YukinoriKAKAZU,” Macroscopic Quantitative Observation of Multi-Robot Behavior”,ICCIMA,2001