413

学習オートマトンによる協調動作への自己組織化に関する研究

北大工　○ 久保正男　三上貞芳　嘉数侑昇

要旨

エージェント群が動的に変化する環境でタスクを効率よく処理する為に具体化するべき機能の一つとして協調動作が挙げられる．本論文ではエージェントに学習オートマトンを適用し，自己組織化を行いながら，動的な環境，" Bamboo Taking Story Problem" (BTS)への適応を試みる．

1，はじめに

多数のエージェントの動的環境でのプラン生成は多くの困難な問題を抱えている．プランの生成や補修に必要な計算量がエージェントの数の増加に従い莫大な量になることは，環境への追従を困難なものにする．近年，このような集中管理的なアプローチの困難性が指摘されるにつれ，分散AI等の分野から自律的なエージェントによる自己組織化によるアプローチが注目されている．この手法は個々のエージェントがそれぞれに環境の観測を行い，これに応じて適切な行動を実行することによって全体として良い結果を生むことを目標としている．さらにこの手法が実現することによって，上記のような集中管理的なプラン生成につ付きまとう問題を和らげることも期待されている．[BAMD92]

一般に，自律的なエージェント群を自己組織化するためには環境を如何に観測し，適切な行動を選び出すことが肝要となる．さらに動的な環境では，行動を選び出す為の基準が変化すると考えられ，この基準を獲得することも大きな問題とされている．[MIKA92]

これらの問題を解決する為の手法として学習オートマトンが注目されている．学習オートマトンは環境を観測した結果を入力とし，これに対応する行動を遷移マトリクスに従って選び，再び環境に出力する．次に，環境を観測したときに出力した行動が適切であってならば，選び出す頻度を上げ，逆の場合には頻度を下げる．このような運動をすることよって上記の変化する基準の獲得が期待されている．

従来，このように基準が変化する問題を扱う分野としてゲーム理論が挙げられるがその中でも’繰り返し囚人のジレンマ’が有名である．繰り返し囚人のジレンマは囚人のジレンマを繰り返し行い総合での利得を評価するゲームであり，幾つかの戦略が知られている．[NISI86]代表的な戦略としてしっぺ返し戦略と裏切り戦略が挙げられるが，どちらの戦略も如何にいつ裏切るかが鍵とされている．これは如何に環境に適応するかといった自己組織化への問題と同一視することが出来る．

そこでエージェント群間での’繰り返し囚人のジレンマ’を実現する問題として’Bamboo taking story '(BTS)　問題を提案する．

2，Bamboo Taking Story 問題　(BTS)

BTSは奇数本の竹の棒をエージェント群で取り合うゲームである．(Fig1)棒は自陣まで引いてはじめてカウントする．一本の棒に多くのエージェントをさく（１），一本当たりのエージェントを少なくする（２）という例を挙げる．(Fig2)どちらの戦略も常に有効でないことがわかる．棒に割くエージェントの割合はゲームの状況に依存する．戦況に応じてエージェントは他のエージェントを助けるために移動を行わねばならない．このような協調動作を実現するために自己組織化を行う必要がある．

3，計算機実験

実験では各エージェントに一つの学習オートマトンを割り当てた．学習オートマトンには確率的学習オートマトン[NARE89]を採用している．あらかじめ与えられた推論方式に従って，取れる棒の数，取られる棒の数，それに必要な時間，誰も触っていない棒の数を計算，それぞれを分割したものと現在の自分のプランを組み合わせ432個の状態を戦況として学習オートマトンへの状態入力とした．各状態に対してオートマトンはどのカテゴリの棒に行くべきか行動を選択する確率を持っている．カテゴリには Free , Pulling, Pulled Adhesion , Pull の５つを用意した．カテゴリが出力されると最短時間で到達できる棒を計算する．評価はプランが達成

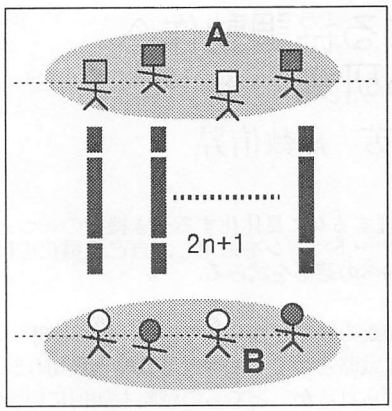

fig.1

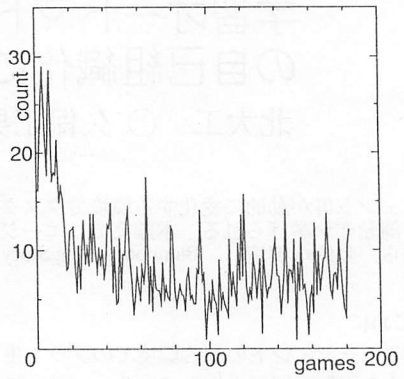

Fig3 convergence;penalty

	場にある棒が取られる確率	引いている棒を奪われる確率	陣地まで引く為に必要な時間
(1)	HIGH	Low	short
(2)	Low	HIGH	long

fig.2 利得に関する一例

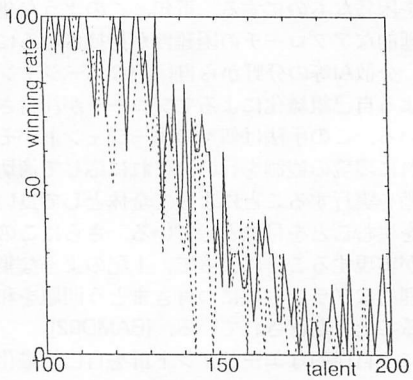

Fig.4;adaptation

されるか，プランが放棄されたときに行われる．評価基準はプラン生成時の戦況との比較で行った．この実験では棒の数を7，棒の長さを8.0，棒から陣地までの距離を30.0，棒間の距離を10.0とした．またエージェントの数を一群4とし標準移動距離を4.0，標準牽引力を3.0と設定した．各エージェントは自陣のランダムな位置よりスタートさせた．

Fig3では学習の収束性について示している．学習回数が進むにつれペナルティの与えられる回数が減り環境への適応が進んでいると考えられる．Fig4は環境への適応に関するグラフである．この実験では100ゲーム同能力の相手と相互に学習後，ランダムにプランを選択する相手と対戦させた．相手は1ゲームごとに標準能力が1%づつ上昇する．縦軸に勝利するパーセンテージをとり，横軸に相手の能力をパーセンテージで表示し学習の有/無による比較を行った．点線は学習無し，実線は学習有りを示している．学習を導入したエージェント群の勝率の落ち込みが緩やかなのがわかった．戦況は相手の能力に応じて変化して行く訳であるが学習によって，勝つための協調動作を実現するために自己組織化を行っていると考えられる．

4．おわりに

動的環境における協調動作実現を目的として学習オートマトンによる自己組織化BTS問題を提案し，実験により有効性を確認した．

参考文献

[BAND92] 坂東友則　確率的学習オートマトンを用いた倒立振子の制御　；卒業論文
[MIKA92] 三上，嘉数　機構論1992
[NARE89]Narendra,Thathachar Learning Automata an introduction
[NISI86] 西山賢一　勝つためのゲーム理論；適応戦略とはなにか；ブルーバックス