3G1-03

エージェントの行動を予測するための意識モデル

Conscious model for predicting agent's action

上妻 将文[*1] 瀧 寛和[*1] 松田 憲幸[*1] 安部 憲広[*2] 堀 聡[*3]
Masafumi Kozuma Hirokazu Taki Noriyuki Matsuda Norihiro Abe Satoshi Hori

[*1] 和歌山大学システム工学研究科 [*2] 九州工業大学情報工学部 [*3] ものつくり大学
[*1]Graduate School of Systems Engineering, Wakayama University
[*2]Faculty of Computer Science and Systems Engineering, Kyushu Institute of Technology
[*3]Institute of Technologists

Abstract: This paper describes the agent system which predicts other agent actions using the reinforce learning. We have been developing the agent which determines actions being conscious of the effect of action. In the system, the agent not only constructs its self-conscious action model, but also learns the model of other agent actions. We also consider the effective action determination method of agents.

1. はじめに

協調して仕事を行うマルチエージェント環境やゲーム環境では，複数のエージェント間の行動には多くの相互作用がある．エージェントが協調して仕事を行うためには，パートナーとなるエージェントの行動を予測できる必要がある．また，他のエージェントの行動を予測することでゲームを有利に進めることができる．他者の行動を予測して適切な行動を選択する仕組みは，知能ロボットやエージェントシステムと人間のインタラクションにおいても，人の意図を把握した親切な行動生成にも必要である．不完全情報下でのゲーム環境では，それぞれのエージェントの戦術の組み合わせでエージェントの利得が定まるものである．

本研究では，自己の行動決定モデルを他者の意識モデルと仮定する．戦術テーブル中の自己の行動（戦術）と他者の行動の利得はエージェントの意思決定に使用する．エージェントの意識モデルをモデル化する方法として，事例ベースによる新戦術の獲得と，強化学習による各エージェントの戦術の組み合わせによる利得の学習と，帰納的学習による意思決定ルールの獲得の方法を提案する．人とエージェントの対話にも，相手の意図理解が重要である．この成果は，利用者の行動から意図を推測できる効率的・効果的な Human-Computer インタラクションを実現できる．

2. システム概要

2.1 エージェントの関係

エージェントは意識モデルを用いて他のエージェントの行動を推測する．エージェントは観測した環境と意識モデルが推測した行動から自己の行動を決定する．他者が実際に選択した行動と意識モデルが推測した行動とを比較してモデルの修正を行う図 1）．エージェントAG1 はエージェントAG2 のモデルからAG2 の行動を推測する．行動システムは，推測された行動および環境条件から自己の行動を決定する．観測システムは AG2 が実際に選択した行動を観測しAG2 のモデルが推測した行動を比較してモデルの修正を行う

連絡先：〒640-8510，和歌山県和歌山市栄谷 930，
和歌山大学システム工学研究科
Tel:073-457-8122，e-mail:s041043@sys.wakayama-u.ac.jp

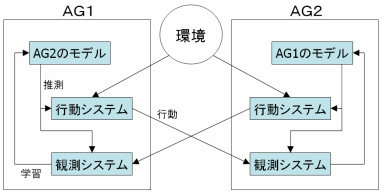

図 1 システムの概念図

2.2 エージェントの戦術表現

エージェントの戦術は，環境条件と選択可能な戦術と成功利得の組から成る戦術テーブルで表現する．また，各戦術はエージェントの行動オペレータの系列で表す．戦術テーブルは全ての環境毎に定義される．(表 1) AG1 の行は AG1 の選択することができる戦術を示す．同様に, AG2 の列は AG2 の選択することができる戦術を示す．例えば，表1の中の AG1 の ST1 とAG2 の ST2 の示す値(4,10)は，AG1 が ST1 を実行しAG2 が ST2 を実行した場合のそれぞれの成功利得を表す．

表 1 エージェントの戦術テーブル

AG2 \ AG1	ST1	ST2	ST3
ST1	10，2	9，2	7，5
ST2	4，10	3，7	3，8
ST3	6，4	6，6	4，7

2.3 戦術の選択

戦術の選択方法について説明する．まず，環境を観測し使用する戦術テーブルを選択する．次に，戦術テーブルから行動を選択する方法としてゲーム理論のナッシュ均衡を用いる．ナッシュ均衡とは，互いに相手の選択した戦術に対して最適な戦術を選択する行動である．ナッシュ均衡を用いることで自己の利得が最大になる戦術を選ぶことが期待できる．表1で, AG1 の戦術 ST1 は AG2 の戦術 ST2 に対する最適反応であり，AG2 の戦術 ST2 は AG1 の戦術 ST1 に対する最適反応である．したがって, AG1 の戦術 ST1 とAG2 の戦術 ST2 の組み合わせがナッシュ均衡点である．

2.4 戦術テーブルの学習

戦術テーブルの成功利得の学習には強化学習を用いる．強化学習は試行錯誤によって未知の環境に適応させる学習制御

の枠組みである．エージェントは環境を観測し，その環境に応じて行動する．行動の結果に報酬を与えることで実行した行動の学習をする．本システムでは戦術テーブルの成功利得が学習される．表1で AG1 が戦術 ST3 を選択し AG2 が戦術 ST2 を選択する場合, AG1 の成功利得は AG2 より低いので AG1 の戦術 ST3 の報酬は低く与えられる．成功利得の学習は以下の式によって行う．

Profit of Agx (AGx-STn, AGy-STm)
= Profit of Agx (AGx-STn, AGy-STm) + Remuneration.

相手より成功利得の高い戦術の時の報酬を 1 とし低い戦術の報酬を-1 とした場合, AG1 が戦術 ST3 を実行し AG2 が戦術 ST1 を実行すると，この組み合わせの成功利得の組は(3,8)から(2.9)に更新される(表 2)．この方式での戦術の学習効果評価[1]は実証されている．

2.5 戦術の追加・削除

戦術テーブルにない未知の戦術事例を観測した場合，新しい戦術として戦術テーブルに追加する．戦術は事例として蓄積する．未知の戦術を次々と戦術テーブルに追加していくと戦術テーブルのサイズが大きくなり戦術の選択に多くの時間が必要になるので，使用頻度の少ない戦術や成功利得の小さな戦術は戦術テーブルから削除する．

表 2 学習後の戦術テーブル

AG2 \ AG1	ST1	ST2	ST3
ST1	10, 2	9, 2	7, 5
ST2	4, 10	3, 7	**2, 9**
ST3	6, 4	6, 6	4, 7

3. 戦略モデル

戦術テーブルは，エージェントの環境における戦術の成功・失敗の可能性を得点化したものである．実際のエージェントの意識は，これらの戦術テーブルの状況をみて戦略を建てることができる．２章でのシステムでは，この戦略をナッシュ均衡で解いた．しかし，意識を持つエージェントを考えると，「裏をかいた行動」「先を予測した行動」「フェイント（ある戦術を選択するそぶりを見せて別の戦術をとる）」など決まりきった戦略をとらない．次の図 2は，戦術テーブルの評価と意思決定のモデルを表現している．

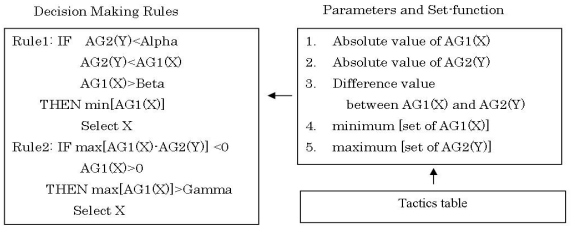

図 2 意思決定モデル

図２において，パラメータと集合関数は，戦術テーブルから行動決定に利用するパラメータと集合の選択機能を示している．重要なパラメータには，エージェント１の戦術 n における成功利得（AG1(ST-n)で表現する）や，エージェントの成功利得の差，戦術集合の中から最小の成功利得を持つ戦術の選択関数 minimum [set of AG1(X)]がある．意思決定ルールはこれらのパラメータを評価して，どの戦術を選択するかを決定する．図２の Rule1 は，相手の戦術の成功利得 AG2(Y)が基準値 Alpha 以下で, AG2(Y)<AG1(X)となるエージェント１の戦術 X が存在し，この成功利得 AG1(X)が基準値 Beta を超えているときに，戦術 X の成功利得のうち最小のものを選択する規則である．

4. 相手の行動を推測する戦略モデルの獲得

相手がナッシュ均衡の戦略で行動を決定していると限らないときには，相手の行動を動的にモデル化する必要性がある．この戦術選択のモデルが、３章で述べた戦略モデルである．戦略モデルでは，ある一瞬の戦術の選択基準と相手の将来の行動を見越した戦術の計画モデルが考えられる．まず，ある一場面での相手の戦術を予測する相手戦略モデルの獲得について述べる．つまり，相手のエージェントの意思決定ルール（図 2参照）を学習することになる．相手の戦術行動を事例として蓄積し，その事例を一般化する帰納的学習によりルールを獲得する．事例の要素は図２のパラメータで表現する．相手行動の予測結果（相手の戦術）と実際の相手の行動を比較することで，相手の行動が予測できたどうかの評価が可能である．予測できた場合には，その相手の戦略ルールのモデルを保持する．予測が外れた場合には，事例に蓄積し帰納的学習を再度適用することでルールを更新する(図 3)．

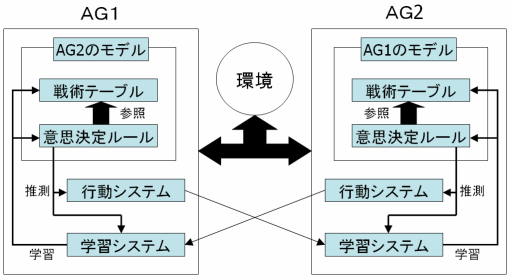

図 3 戦略モデルをもったシステムの概念図

5. まとめ

本研究では，マルチエージェントのための意識モデルおよび不完全情報ゲーム下での意識モデルを用いた戦術選択の手段を提案した．本システムは自己の行動決定モデルを用いて他者の意識モデルを学習する．戦略モデルの実装と予測実験は現在進行中であるが，まだ結果を得ていない．本研究では，計算時間は考慮しない（考慮しなくても良いゲーム）で検討し，将来的には，実時間のゲーム（ロボットサッカーなど）に拡張予定である．

参考文献

[1] 上妻将文，瀧寛和，松田憲幸，安部浩一，堀憲広：強化学習を利用した意識モデルの構築，第 4 回 SICE システムインテグレーション部門 講演会論文集, 2B4-5, 2003
[2] 新田克巳：電子・情報工学講座２４ 人工知能概論, 176，培風館
[3] 作田誠：不完全情報ゲーム研究の現状，情報処理, 44 巻, 9 号, 916-920, 2003
[4] 野田五十樹：お手軽サッカーを目指して，情報処理, 44 巻, 9 号, 928-930, 2003
[5] 福井, 瀧, 松田, 安部, 堀：意識エージェントの構想，人工知能学会全国大会(第 16 回)論文集, 2b3-05, 2002
[6] Richard S. Sutton and Andrew G. Barto: Reinforcement Learning. MIT Press, 1998
[7] R.J. Aumann: Lectures on Game Theory. Westview Press, 1989