人　工　知　能　105－3
（1996．8．26）

相手エージェントを考慮した行動戦略の調整

松原 繁夫　　横尾 真
ＮＴＴコミュニケーション科学研究所

本論文はマルチエージェント環境でエージェント間に計算能力差が存在する場合の効率的な行動選択法を提案する．ここでは，短期的には損をする可能性があるが長期的に大きな利益を得ることが可能な行動を協調行動とする．このとき，計算能力の低いエージェントは長期的損得を計算できず非協調行動を選択する．そのため，結局システム全体が非協調行動に陥り，利得の低い状態で安定してしまう．この問題を解消するため，本研究は獲得利得に関する計算結果を必要に応じて相手に通知し，通知された方はそれを自己の計算結果と合成して行動を選択する方法を提案する．合成時には，相手の計算結果と自己の計算結果の類似性を評価する．これによりエージェント双方にこの方法を採用する動機が生じる．また，虚偽の通知を行い相手を陥れようとする動機の抑制も達成される．例題を用いた評価により，双方が協調行動を取り，総獲得利得の増加を計れるという提案手法の有効性を示す．

Action Selection by Incorporating the Other Agent's View into Self-View

Shigeo Matsubara and Makoto Yokoo
NTT Communication Science Laboratories
2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan

This paper proposes a new efficient method to select an action on condition that heterogeneous agents exist. This method is as follows: a agent with poor computation resources selects an action by incorporating a view of the other agent with rich computation resources into a view of himself. The way of incorporation is based on the similarity between their views, which gives them an incentive to adopt the method and restrain them from cheating. By using our proposed method, their agents are able to avoid falling into a less productive state and reach a cooperative state.

1 はじめに

マルチエージェント環境でエージェントが適切にふるまうためには，他エージェントとの関係を考える必要がある．この種の関係を記述解析する方法としてゲーム理論がある．ゲーム理論は“囚人のジレンマ”の例などでよく知られており，分散人工知能の分野でもゲーム理論の適用が試みられている [Rosenschein and Zlotkin, 1994], [Genesereth *et al.*, 1986].

行動選択を複数回繰り返す繰り返しゲームでは，従来のゲーム理論を用いると，繰り返し数が有限であれば非協調行動しか現れない．これは現実世界での人や組織の行動と一致しない．また，エージェントの獲得利得が低いレベルに留まる．この問題を解消するため，筆者らは行動実行による利得値の変化を導入した [松原, 横尾, 1996], [Matsubara and Yokoo, 1996]．そこでは，短期的には損をする可能性があるが長期的に大きな利益を得ることが可能な行動を協調行動と考え，逆の場合を非協調行動と考えた．長期的な損得を評価することで協調行動が出現する．

さて，エージェントの能力を非均質と仮定すると新たな問題が生じる．エージェントの計算能力に制限がある場合の問題解決は限定合理性という用語を用いて広く研究されている．本研究ではその計算能力にエージェント間で差が存在する場合を扱う．計算能力の低いエージェントは長期的損得を計算できず近視眼的な意思決定を行い，非協調行動を選択する．このため，結局システム全体が非協調行動に陥り，利得の低い状態で安定してしまう．

そこで，本論文では上記の問題を解消して双方の協調行動を可能とする新たな行動選択法を提案する．まず，計算能力の高いエージェントが必要に応じて利得に関する自己の計算結果を相手に通知する．それを受けて計算能力の低いエージェントは自己の計算結果と通知された相手の計算結果を合成し，その上で行動選択を行う．ここで問題となるのは双方がこの方法に従う動機を持つかどうかである．提案方法では，合成時に自己の計算結果と相手の計算結果の類似性を評価する．これにより，たとえ通知側が相手を操作しようとして虚偽の通知を行っても，被通知側の利得が減少する可能性が小さくなる．この事実から，計算能力の低いエージェントも提案方法採用の動機を持つに至る．また，通知側も虚偽の通知を行わないことが獲得利得の増加に結び付くことがわかり，虚偽の通知をする動機を持たなくなる．これらから，双方が協調行動を取って，総獲得利得の増加を計ることが可能となる．例題を用いた評価により提案手法の有効性を示す．

	C_2	D_2
C_1	3 3	4 1
D_1	1 4	2 2

行動組	変化分
(C_1, C_2)	α
(C_1, D_2)	β
(D_1, C_2)	β
(D_1, D_2)	β

C_1, C_2: 協調, D_1, D_2: 非協調

図 1: 利得行列によるゲームの表現

2 ゲーム理論的アプローチとその問題

2.1 ゲーム理論による表現と行動選択法

ゲームはエージェント，エージェントの実行可能な行動，利得の3つの要素からなる．ゲームは図1に示す利得行列を用いて表現される．これはエージェントの世界の見方を示す．図1はエージェント $A1$ と $A2$ がそれぞれ協調行動か非協調行動を選択でき，例えば，$A1$ が協調行動 C_1，$A2$ が非協調行動 D_2 を実行した場合，$A1$ と $A2$ がそれぞれ利得1と4を獲得することを表す．エージェント間での行動選択は1回切りとは限らず，このゲームが何度か繰り返される場合もある．本論文では行動実行により利得行列が変化すると仮定する．各行動組が実行された場合，図1右に示す変化分が利得行列の全成分に加えられるとする．本研究ではさらに以下の仮定をおく．

- 2エージェント間の関係に限定する．3エージェント以上の関係は，2エージェント間の関係の積み重ねとみなす．
- エージェントは合理的である．すなわち，獲得利得最大化を目指す．
- エージェント間に事前に強制力のある取り決めは存在しない (非協力ゲーム)．

	C_2*	D_2*
C_1*	7.4 7.4	8.0 5.0
D_1*	5.0 8.0	6.0 6.0

* はそれ以降のゲームでの評価した範囲内での最適の行動系列を表す.

図 2: 累積利得行列の表現例

- エージェントは相手の利得を知っている. 利得値の変化の仕方も知っている.

つぎに, 上記の表現が与えられたときの行動選択法を述べる. ここでは, 行動選択法として後ろ向き推論法を用いる. 先読み数を n とすると, 後ろ向き推論法では, 各行動系列に対する n 手先の利得行列を計算する. その利得行列上で均衡点を求める. 均衡点に属する行動組を選択行動とする. 均衡点とは, 他のエージェントがある行動組に従うならば, 自己もその行動組に従うことが自己の利得を最大にするという性質を持つ行動組を指す. この均衡点の利得値を $n-1$ 手先の利得行列の対応成分に加え, この上で均衡点を求める. すると, $(n-1) \sim n$ 手での均衡点に属する行動系列が求まる. この計算を 1 手目まで繰り返す.

後ろ向き推論法では, n 手先まで評価した範囲内での最適の戦略が求まる. この方法で求まる, 1 手目から n 手目までの獲得利得の総和を表す利得行列を累積利得行列と呼ぶ (図 2参照). 計算量は先読み数 n の増加にともない, 指数的に増加する.

2.2 エージェントの非均質性による問題

マルチエージェントシステムでは, 全てのエージェントが同等の能力を持つと仮定できない. エージェント間には以下の能力差が存在する.

1. 情報獲得能力の差

   利得行列の利得値の差となって現れる.

2. 推論能力の差

   獲得利得の先読み能力の差として現れる.

3. 実行可能行動の差

   利得行列の実行可能行動の有無, 利得値の差として現れる.

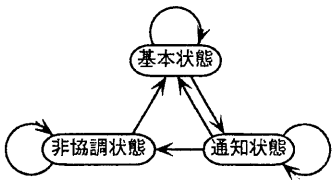

図 3: 利得行列合成による行動選択法

いずれも, どの行動を均衡点に属すると判断するかで差が生じる. 本論文では 2. の推論能力の差のみについて考える. 1 章で述べたように, エージェント間で推論能力が異なれば, 各々別個の行動を均衡点に属すると判断することが起こる. 推論能力の高いエージェントが協調行動を選択しても, 推論能力の低いエージェントが同調して協調行動を取らなければ, 結局は双方が非協調行動に陥る. そうなれば, 低い利得で安定してしまい, より大きな利得獲得の機会を逃すことになる.

3 利得行列合成による行動選択

前章で述べたエージェント間の計算能力差から非協調行動に陥ることを防ぐため, 新たな行動選択法を提案する. 提案方法の要点を以下に示す.

1. 行動実行の各回において獲得利得値が期待値より小さければ, 相手に自己の計算した累積利得行列を通知する. 相手から累積利得行列の通知があれば, それを自己の計算した累積利得行列と合成し, その上で行動を選択する.

2. 自己と相手の計算結果間の類似性を元に, 相手の計算結果を参照する程度を変える.

1. により, 計算能力の高いエージェントの計算結果を計算能力の低いエージェントが用いることで, その能力差を埋めることが可能となる. 2. は計算能力の低いエージェントが相手を信用するために必要な機構である.

行動選択法の詳細を以下に示す. エージェントは図 3 に示す 3 つの状態を遷移する. 各状態内での計算手順を以下に示す.

[利得行列合成による行動選択法]

(a. 基本状態)

a1. 自己の累積利得行列 M_{self} を計算する.

a2. 相手から累積利得行列 $M_{opponent}$ の通知がなければ, M_{self} 上で均衡点を計算する.

a3. 相手から累積利得行列 $M_{opponent}$ の通知があれば, 合成利得行列 M_c を計算し, その上で均衡点を計算する. 合成法は本章後半で示される.

a4. 均衡点に属する行動を実行する.

a5. 行動実行による獲得利得が期待値より小さければ, (a. 通知状態 (l)) へ行く ($l \geq 1$).

a6. a1. へ戻る.

(b. 通知状態 (l))

b1. 相手に自己の累積利得行列 M_{self} を通知する.

b2. (a. 基本状態) の a1. ~ a4. を実行する.

b3. 行動実行による獲得利得が期待値より小さければ, (c. 非協調状態 (m)) へ行く ($m \geq 1$).

b4. $l \leftarrow l-1$ とし, $l=0$ であれば, (a. 基本状態) へ戻る. $l=0$ でなければ, b1. へ戻る.

(c. 非協調状態 (m))

c1. 非協調行動を実行する.

c2. $m \leftarrow m-1$ とし, $m=0$ であれば, (a. 基本状態) へ戻る. $m=0$ でなければ, c1. へ戻る.

累積利得行列の合成法を以下に示す. 先読みによる累積利得行列の変化に類似性があれば, 相手の累積利得行列をより重視して行動選択を行う.

[利得値の変化の類似度に基づく合成法]

1. 自己の計算した累積利得行列 M_{self} を最小値が0, 最大値が1 となるよう線形変換する. 変換後の利得行列を標準化累積利得行列 M_{self}^{normal} と呼ぶ.
2. 相手の計算した累積利得行列 $M_{opponent}$ を最小値が0, 最大値が1となるよう線形変換する. 変換後の利得行列を標準化累積利得行列 $M_{opponent}^{normal}$ と呼ぶ.
3. 以下の式に基づいて合成計算を行う.

$$\text{合成利得行列 } M_c = (1-w) \times M_{self}^{normal} + w \times M_{opponent}^{normal}$$

ここで,

$$\text{合成重み } w = \frac{(\vec{s}, \vec{o})}{\|\vec{s}\|\|\vec{o}\|} \qquad (3.1)$$

$\vec{s}$: $(M_{self}^{normal} - M_{init}^{normal})$の列ベクトル表現
$\vec{o}$: $(M_{opponent}^{normal} - M_{init}^{normal})$の列ベクトル表現
M_{init}^{normal} : 初期状態の標準化利得行列
$(\vec{s}, \vec{o})$: $\vec{s}$と$\vec{o}$の内積, $\|\vec{s}\|(\|\vec{o}\|)$: $\vec{s}(\vec{o})$のノルム

4 例題を用いた行動選択法の評価

4.1 利得行列合成の効果

本章では利得行列合成による行動選択法の性質を調べる. まず, 利得行列合成時の重みと選択行動の関係を調べる. 高計算能力エージェント A_H と低計算能力エージェント A_L が図1 に示すゲームを繰り返し行うとする. 変化分 $\beta = 0.0$ で固定とし, 変化分 $\alpha = 0.1, 0.2, 0.3$ の各場合について調べる. エージェント A_H は $n(>3)$ 手先まで先読みして行動選択し, エージェント A_L は3手先まで先読みして行動選択すると仮定する. A_L は単独では非協調行動を選択する.

図4に利得行列合成による行動選択法を用いた場合の A_L の選択行動を示す. ただし, ここでは行動選択法の性質を調べるため重み w を変化させている. (a), (b), (c) それぞれ, $\alpha = 0.1, 0.2, 0.3$ の場合に対応する. 各グラフの横軸は A_H の先読み数 n, 縦軸は合成時の重み w を示す. 図4で陰のつけてある部分は, A_L が協調行動を選択する領域を示し, 陰のない部分は, A_L が非協調行動を選択する領域を示す. 図4から以下のことがわかる.

- A_L に協調行動を取らせるには, 重み w をできるだけ大きく設定するとよい.
- A_L に協調行動を取らせるには, A_H の先読み数をできるだけ増やせばよい.
- 協調行動による利得値の増分が大きくなる (α が大きい) と A_L が協調行動を取る可能性が大きくなる.

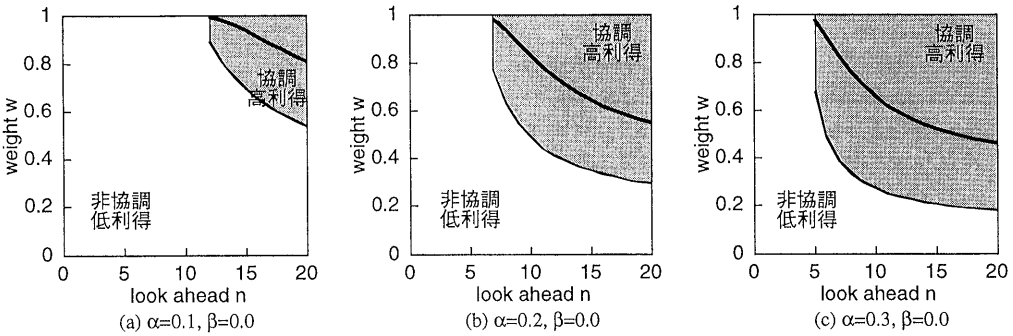

図 4: 合成重みと選択行動の関係

	C_L*	D_L*
C_H*	0.87 0.87	1.0 0.0
D_H*	x 1.0	0.33 0.33

図 5: 虚偽の利得行列

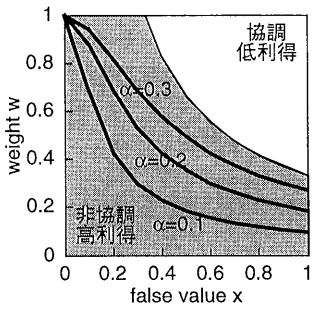

図 6: 虚偽通知の選択行動への影響

4.2 虚偽内容通知による影響

前節で見たように, A_L に協調行動を取らせるには, 重み w を大きくすればよい. しかし, 重み w を大きくすることは, A_L が A_H に操作されやすくなることを意味する. この確認のため, A_H が偽って図 5に示す利得行列を A_L に通知すると仮定する. この利得行列中の x は A_H が非協調行動 D_H を取っても, A_L が協調行動 C_L を取ろうとする程度, つまり, A_H が自己の有利な方向へ誘導する程度を表す. $x = 0$ が真値であり, x が大きいほど虚偽の程度が大きくなる. x 以外の値は, 3 手先読みした場合の標準化累積利得行列の値である.

図 6に A_L の選択行動を示す. 横軸は虚偽の程度 x を表し, 縦軸は合成時の重み w を表す. 図 6で陰をつけてある部分は A_L が A_H に操作されずに非協調行動を取る領域を表す. 陰のない部分は A_H が非協調にもかかわらず, A_L が騙されて協調行動を取る領域を表す[*1].

[*1] A_H が非協調行動を取るため, A_L は非協調行動を取る方が自己の利得が大きくなる.

図 6から以下のことがわかる.

- 重み w を大きく設定すると A_L は A_H の操作を受けやすくなる.
- 虚偽の程度が大きくなると A_L は A_H の操作をより受けやすくなる.

よって, w が一定値に設定される場合, 以下の理由から A_L は利得行列合成による行動選択法採用の動機を持たない.

- 重み w を小さく設定すると, 利得行列合成の効果がなく, A_L は協調行動を取れない. そのため長期的利得値が増加しない.
- 重み w を大きく設定すると, A_L は A_H の操作を受けやすくなり, 利得値が低下する危険性がある.

4.3 合成に類似度を用いる効果

本節では利得値の変化の類似度による重み w の決定が, A_L に提案方法採用の動機を与えることを示す. 利得値の変化の類似度 (式 (3.1)) から計算した重み w の値を図 4と図 6に太線で示してある. これらから以下のことがわかる.

- A_H が正しい利得行列を通知する場合, w の値は A_L が協調行動を取る領域に入る.
- A_H が偽った利得行列を通知する場合, w の値は A_L が非協調行動を取る領域に入る[*2].

すなわち, 類似度を用いれば, A_L の利得が減少する可能性が低下する. よって, 類似度を用いて w を設定すれば A_L に提案方法採用の動機が生まれる. また, 偽った利得行列を通知すれば, A_L が非協調行動を取る, すなわち, A_H の獲得利得が増加しないことがわかるため, A_H には正しい利得行列を通知する動機が生まれる.

4.4 計算結果通知の方向性

利得行列合成による行動選択法では, 計算能力の高い側から低い側へ計算結果の通知が行われると双方の協調行動に至る. しかし, 低い側から高い側へ通知が行われると全体が非協調行動に陥ることが起こる. 本論文では獲得利得が期待値より低い場合にのみ通知を行うことで, 低い側から高い側への通知の排除を試みた. 別の対処法としては, 計算結果の通知に要するコストに見合う利得の増加があるかどうかを評価し, それをその後の通知／非通知の判断に反映させることが考えられる. 詳細な検討は今後の課題である.

4.5 総獲得利得の比較

図 1に示すゲームにおいて, 変化分 $\alpha = 0.2, \beta = 0.0$, 高計算能力エージェント A_H の先読み数 10, 低計算能力エージェント A_L の先読み数 3 として他の条件を変えた場合の, 繰り返し数 10 回目までの各エージェントの総獲得利得を表 1 に示す. 提案手法を用いることで, 長期的に大きな利得を獲得できることがわかる.

[*2] $\alpha - \beta$ の値が大きくなると w の値が A_L の協調領域に入るようになる. しかし, A_L は単独で最初から協調行動を選択でき, 利得行列の通知が起こらないためその影響は小さい.

実験条件 ($l = 5, m = \infty$)			獲得利得	
通知	虚偽	重み w	A_H	A_L
無	–	–	18.0	24.0
有	無	類似度による	32.6	38.6
有	無	固定 ($w = 0.8$)	32.6	38.6
有	有 ($x = 0.6$)	類似度による	18.0	24.0
有	有 ($x = 0.6$)	固定 ($w = 0.8$)	21.0	24.0

表 1: 総獲得利得の比較

5 むすび

本論文では, 利得に関する自己の計算結果を必要に応じて通知し, 通知がある場合には自己と相手の計算結果を合成して行動選択する方法を提案した. この方法はエージェント間に計算能力差が存在する場合に, 双方が非協調行動を取って獲得利得の低い状態に陥ることを回避できる. 計算結果の合成時に, 自己と相手の計算結果の類似度を利用することで, 双方に提案方法を採用する動機を持たせ, また, 虚偽のない通知を行う動機を持たせることが可能となった. 提案方法を用いれば, 双方が協調行動を取って長期的な獲得利得の増加を計ることが可能となる. 本論文では, 行動による利得の変化を一様と仮定した. より複雑な変化パターンに対し, 提案方法の性質を調べることは今後の課題である.

最後に, 本研究を支援して頂いた NTT コミュニケーション科学研究所松田晃一所長, 大里延康主幹研究員に感謝します.

参考文献

[Genesereth *et al.*, 1986] Genesereth, M. R., Ginsberg, M. L., and Rosenschein, J. S., "Cooperation without Communication," *AAAI-86*, pp.51-57, 1986.

[松原, 横尾, 1996] 松原, 横尾, "多段階ゲームとして見たエージェントの行動選択," 1996 年度人工知能学会全国大会 (第 10 回) 予稿集, pp.159-162, 1996.

[Matsubara and Yokoo, 1996] Matsubara, S. and Yokoo, M., "Cooperative Behavior in an Iterated Game with a Change of the Payoff Value," *ICMAS-96*, 1996. *(to appear)*

[Rosenschein and Zlotkin, 1994] Rosenschein, J.S., and Zlotkin, G., *Rules of Encounter*, The MIT Press, 1994.