社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report

2006－ICS－142 (7)
2006／1／12

情報操作による社会の安定化

太田正幸† 野田五十樹†

利己的なエージェントの集まりによるマルチエージェント環境において、全エージェントが大域情報にアクセス可能な場合、それを用いるとかえってエージェントの政策が不安定化し、全体の効用の合計が低下する可能性がある。この問題に対し、我々は大域情報を多少加工して流すことによりエージェントの政策をNash均衡に収束、安定化させ、全体の効用の合計が低下することを防ぐ方法を提案する。この手法には、全体のエージェント数やNash均衡時の人数配分といった、問題依存の情報が不要であるという利点がある。本稿では、資源配分問題を用いて安定性、収束性の2つを示し、シミュレーション実験により提案手法の有効性を確認した。

A Method to Reduce Bad Influence of Information Sharing

MASAYUKI OHTA† and ITSUKI NODA†

In a multi-agent environment that consists of selfish agents, the global-information has possibility to have negative effects to the agents. It may happen because all agents show tendency to take the identical alternative by considering the information and the utility of each agent decreases as a result. To overcome this problem, we propose a method to converge agents' strategies to Nash equilibrium and to stabilize there, by releasing modified global-information. This method also prevent from decreasing the total utility of agents. It is advantageous in that information which depends on the problem, such as total number of agents and distribution at Nash equilibrium, is not necessary. The stability and convergency of the proposed method are proved using a resource allocation problem, and the effectiveness is confirmed by simulation experiments.

1. はじめに

近年、全てのエージェントが自分の利益だけを考えて行動する(利己的である)ようなマルチエージェント環境において、全体の効用合計を可能な限り高い値に保つための制御方法に注目が集まっている[8]。このような制御を必要とする状況は現実の社会の様々な場面に存在し、交通制御[4),7)]や人工市場[3)]といったの分野への適用が始まっている。

我々は、エージェントの知覚能力に制限がある場合を想定し、自分で獲得した情報(局所情報)だけを用いて行動ルール(政策)を更新する場合と、他のエージェントが獲得した情報(大域情報)も全て用いることができる場合との違いに着目した。個々のエージェントが、全体の効用合計を最大化することを目的としている(協調的である)場合には、大域情報は全体の効用合計を向上させることに大きく寄与することを示した研究が数多く存在する[5),6)]。しかし、エージェントが利己的である場合には、個々のエージェントが自分の利益だけを追求するため、仮にある選択肢を取ることにより非常に高い効用が得られるという大域情報が流れれば、全てのエージェントがその選択肢を取るようになり、結果的に、その選択肢を取ることにより得られる効用は大きく減少する可能性がある。また、長い目で見ても、上記の現象によりエージェントの行動および得られる効用が不安定になることは、あまり好ましい状況ではないことが多い。この問題に対し、我々はエージェント間で伝わる大域情報をそのまま流すのではなく、多少加工した上で流すことにより全体を安定化させる手法を提案する。

2. 問題設定

我々は、前節で述べたような大域情報の悪影響について、資源配分問題を用いて議論する。この節では、本稿で用いる資源配分問題の定式化を行なったのち、上記、悪影響の原因について整理する。本稿の目的は、個々のエージェントの効用を最大化する政策を求めることではなく、エージェントが(固定の戦略に基づき)利己的に行動することを前提に、流れる大域情報を加

† 産業技術総合研究所
National Institute of Advanced Industrial Science and Technology (AIST)

工することで、全体の政策を安定化させ、結果的に全体の効用の合計が低下するのを防ぐことにある。

2.1 資源配分問題

本稿では、大域情報の悪影響について議論するため以下のような資源配分問題を考える。まず、環境には s 個の資源 $R = \{r_1, r_2, \cdots, r_s\}$ と t 人のエージェント $A = \{a_1, a_2, \cdots, a_t\}$ が存在するものとする (図 1 参照)。エージェントの目的は、資源を選択することによって得られる効用の期待値の最大化である。各エージェント a_i は、個別にそれぞれの資源 r_j に対する効用の推定値 $V_i(r_j)$ を保持しており、各離散ステップ毎に、その値に従って 1 つの資源を選択する。資源 r_j を選択した場合にエージェント a_i が得る効用は、あらかじめ用意された効用関数 $U_j(n_j)$ (ただし、n_j は r_j を選択したエージェント数) により、同じ資源を選択したエージェント数に基づいて決定される。効用が観測されると、エージェントは自分で観測した効用および何らかの方法で得た大域情報に基づき、効用の推定値を更新し、次のステップへと移行する。以上の操作を、全エージェントの政策が安定するか、もしくは安定しないことが分かるのに充分な回数繰り返す。

同様の評価環境として Minority Game[2] や N 人版囚人のジレンマゲーム[1] などが存在するが、この資源配分問題は行動の選択肢が 3 つ以上存在するという点でそれらとは異なる。

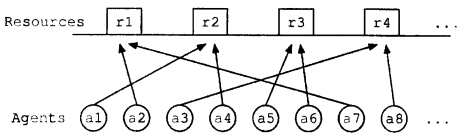

図 1 資源配分問題

2.2 エージェント

本稿で扱う資源配分問題では、エージェントは常に $V_i(r_j)$ が最大の資源を選択 (グリーディーに行動) する。また、問題を単純化するため、探査的な行動は一切取らないものとする。従って、エージェントは自分が選択しなかった資源に関しては大域情報によってのみ情報を得ることができる。得られた効用の情報 $U_j(n_j)$ に対し、エージェントは学習率を $\alpha(0 < \alpha \leq 1)$ として、次の式で効用の推定値を更新する。

$$V_i(r_j) = (1-\alpha)V_i(r_j) + \alpha U_j(n_j)$$

この更新は、局所情報、大域情報に関わらず、受け取った全ての情報に対して 1 回ずつ行なわれる。

2.3 大域情報

大域情報は、実世界ではネット上の情報、メディアが流す情報，噂などに相当する。本稿では、他のエージェントが観測した効用の値がそのまま大域情報として流れることを考える。従って、誰も選択しなかった資源の効用に関する情報は一切手に入らない。この情報は全てのエージェントが同じ値を受け取るため、前述した悪影響が生じ得る。本稿の目的は、各エージェントレベルでの調整ではなく、この大域情報を加工して流し、エージェントの政策変更を間接的に促すことにより全体の動作を制御することにある。もし、効用関数と全エージェント数が分かっているなら、Nash 均衡時のエージェント分布を求めることが可能であるため、その際の効用を大域情報として流すこと等も考えられる。これについては、第 4 節において、実験的に検証を行なう。

2.4 問題点の整理

この資源配分問題において、大域情報が悪影響となる状況を確認する。あるステップにおいて、エージェント a_i $(i = 1, 2, \cdots, t)$ がそれぞれ独自の推定値に基づいて資源を選択した結果、r_j に対する効用が最大であったとする。上記の枠組では、全てのエージェント a_i が等しく r_j に対する個々の効用の推定値を最大の効用を示す $U_j(n_j)$ に基づき $V_i(r_j) \leftarrow U_j(n_j)$ という方向に更新するため、いずれは一部のエージェントが r_j に選択を変更することになる。このため、r_j を選択するエージェント数は r_j で得られる効用が最大でなくなるまで必ず増加し続ける。そして、r_j で得られる効用が最大ではなくなると、また別の、そのとき最大の効用が得られる資源を選択するエージェントが増加し始める。以上を繰り返すため、エージェントの政策は安定せず、Nash 均衡からある程度ずれた分布が定常的に続くため、全体の効用合計もしばしば不安定となり、平均すると、少なくとも Nash 均衡での状態よりは悪い値を示す。学習のパラメタにも依存するが、Nash 均衡付近ではそれぞれの資源に対して得られる効用の値が拮抗しているため、このエージェントの選択の分布が劇的に変化することも少なくない。この問題は全員が同じ情報に基づいて効用の推定値を更新することが原因であり、その点が改善されない限り解決することはできない。

3. 情報操作による安定化

この節では、エージェント間で伝わる大域情報を加工した上で流すことにより第 1 節で述べた大域情報による悪影響を軽減する手法を提案し、その Nash 均衡における安定性，Nash 均衡への収束性を示す。また、Nash 均衡が既知の場合、その情報により提案手法と

類似したアプローチが有効に働くかどうかを検証する。

3.1 Next-To 情報

第 2.4 節で述べたように、大域情報の悪影響は、全く同じ情報が全てのエージェントに伝わってしまうことから生じる。大域情報は実際にいずれかのエージェントが観測した効用であり、その値に基づいて得られる効用の推定値を計算すること自体には問題は無い。しかし、その推定値を使って次の行動を決定することを考えると、自分自身が政策を変更することによって得られる効用の値が変化することが考慮されていない点が大きな問題となる。すなわち、資源 r_j を選択しているエージェント a_i が大域情報として r_k $(j \neq k)$ において $U_k(n_k)$ の効用が得られるという情報を得たとしても、実際に a_i が選択する資源を r_k に変更した場合、他のエージェントの選択に変化が無ければ、r_k の選択人数は n_k+1 となり、得られる効用は $U_k(n_k+1)$ となるため、実際に手に入る効用は $U_k(n_k)$ よりも少ないのである。そこで、我々は、大域情報に関してはエージェントに $U_i(n_k)$ をそのまま伝えるのではなく、そのエージェントが行動を変更した場合のことを考慮し、$U_i(n_k+1)$ に加工して伝えるという方法を提案する。この方法に従うと、r_k で得られる効用を自分で観測したエージェントは局所情報として $U_i(n)$ を用いるため、全員が同じ情報に基づいて効用の推定値を更新するという状況は回避することができる。この方法では、大域情報として実際の効用とは異なる値を流すことになるが、これは、各エージェントが政策を変更した場合に受け取る効用の予測値をそれぞれに対して個別に伝えていることに相当する。

3.2 提案手法の妥当性

上記の提案手法の妥当性を示すため、以下において、Nash 均衡における安定性および、Nash 均衡への収束性について述べる。Nash 均衡は必ずしも最適解である保証は無いが、少なくとも局所的には良い解であることを保証することが可能になる。

3.2.1 安 定 性

現在、全エージェントの資源選択の分布が Nash 均衡の状態にあると仮定し、そのとき各資源 r_j を選択している人数を $\hat{n_j}$ とする。このとき、資源 r_a, r_b を選択したエージェントはそれぞれ局所情報として $U_a(\hat{n_a})$, $U_b(\hat{n_b})$ の効用を得る。あるエージェント a_1 が r_a を選択していたとすると、提案手法を用いた場合、r_a に対する局所情報として $U_a(\hat{n_a})$、r_b に対する大域情報として $U_b(\hat{n_b}+1)$ を受け取ることになる (表 1 参照)。このとき、$\hat{n_a}, \hat{n_b}$ はナッシュ均衡時の人数であり、常に以下の式が成り立つため、どのエージェントも現在の政策を変更する動機を持ち得ない。

$$U_a(\hat{n_a}) > U_b(\hat{n_b}+1)$$

従って、提案手法を用いた場合、少なくとも Nash 均衡では全エージェントの政策は安定することが分かる。

表 1 Nash 均衡における効用の比較 (提案手法)

	r_a	r_a
現在の値	$U_a(\hat{n_a})$	$U_b(\hat{n_b})$
大域情報	$U_a(\hat{n_a}+1)$	$U_b(\hat{n_b}+1)$

3.2.2 収 束 性

現在、エージェントの資源選択の分布が Nash 均衡の状態ではないと仮定すると、選択人数が $\hat{n_a}+x$ (x は 1 以上の整数) の資源 r_a と、選択人数が $\hat{n_b}-y$ (y は 1 以上 x 以下の整数) の資源 r_b とが必ず存在する。あるエージェント a_1 が r_a を選択していたとすると、提案手法を用いた場合、r_a に対する局所情報として $U_a(\hat{n_a}+x)$、r_b に対する大域情報として $U_b(\hat{n_b}-y+1)$ を受け取ることになる (表 2 参照)。このとき、以下の式が常に成り立つため、Nash 均衡時よりも多くのエージェントが選択している資源から Nash 均衡時よりも少ない数のエージェントが選択している資源を選択するように政策の変更が起こる。

$$U_a(\hat{n_a}+x) < U_b(\hat{n_b}) \leq U_b(\hat{n_b}-y+1)$$

一方、別のエージェント a_2 が r_b を選択していたとすると、提案手法を用いた場合、r_b に対する局所情報として $U_b(\hat{n_b}-y)$、r_a に対する情報として $U_a(\hat{n_a}+x+1)$ を受け取ることになるが、この場合には以下の式が常に成り立つため、現在の政策を変更する動機を持ち得ない。

$$U_a(\hat{n_a}+x+1) < U_b(\hat{n_b}) < U_b(\hat{n_b}-y)$$

以上より、提案手法を用いた場合の Nash 均衡への収束性が確認された。

表 2 Nash 均衡付近における効用の比較 (提案手法)

	r_a	r_a
現在の値	$U_a(\hat{n_a}+x)$	$U_b(\hat{n_b}-y)$
大域情報	$U_a(\hat{n_a}+x+1)$	$U_b(\hat{n_b}-y+1)$

3.3 Nash 均衡が既知の場合

Nash 均衡の値が分かっている場合について、上記同様の考察を行なう。Nash 均衡の値が分かっていても、その際の効用を全エージェントに対し大域情報として伝えてしまうと、全てのエージェントに同じ情報が伝わるため、大域情報の悪影響はこれまでの議論と同様、避けることができない。これに対し、提案手法

からの類推として、資源 r_j の Nash 均衡時における効用が $U_j(\hat{n_j})$ である場合、$U_j(\hat{n_j}+1)$ を大域情報として流す手法の妥当性について考える。Nash 均衡の状態にある時には、提案手法の場合と同様 (表 1 参照)、局所情報 $U_j(\hat{n_j})$ と大域情報 $U_j(\hat{n_j})+1$ との間には常に以下の式が成り立つため、Nash 均衡における安定性は保証される。

$$U_a(\hat{n_a}) > U_b(\hat{n_b}+1)$$

一方、Nash 均衡への収束性については、局所情報 $U_a(\hat{n_a}+x)$ と大域情報 $U_a(\hat{n_a})+1$ との間に、$x=1$ のケースが存在するため、以下の関係は保証されない。

$$U_a(\hat{n_a}+x) < U_b(\hat{n_b})+1$$

ただし、$x \geq 2$ の場合については上記の式が成り立つため、大域情報として $U_j(\hat{n_j}+1)$ を用いた場合には、Nash 均衡の一歩手前までの収束のみが保証されるということになる (表 3 参照)。

表 3 Nash 均衡付近における効用の比較 (Nash 均衡情報を使用)

	r_a	r_a
現在の値	$U_a(\hat{n_a}+x)$	$U_b(\hat{n_b}-y)$
大域情報	$U_a(\hat{n_a}+1)$	$U_b(\hat{n_b}+1)$

4. 実　　験

提案手法の有効性を確認するため、2.1 節で示した資源配分問題において、大域情報として以下の 4 通りの情報を流した場合について比較する。

- 他のエージェントが実際に観測した効用 (EX)
- 提案手法により上記の効用を加工した値 (EX+1)
- Nash 均衡時の効用 (EQ)
- Nash 均衡時の人数+1 の場合の効用 (EQ+1)

4.1 実験の詳細

実験における詳細の設定は以下の通りである。まず、資源が 5 つ、エージェントが 100 人存在し、以下の効用関数に基づく環境を考える。

$$U_j(n_j) = u_j/\sqrt{n_j}$$

各資源 r_i に対し、その資源をエージェントが 1 人だけ選択した場合の効用はそれぞれ $u_j = \{10.0,\ 8.0,\ 6.4,\ 6,4,\ 8.0\}$ とする。この環境では、Nash 均衡時のエージェント配分は順に 32, 21, 13, 13, 21 となる。$V_i(r_j)$ は全て 0.0 - 1.0 の間でランダムな値に初期化する。エージェントの学習率 α は、どの情報に対しても全て 0.1 とした。

4.2 実験結果

それぞれの大域情報を用いた場合の学習曲線を図 2 に、その際のエージェントの分布状況を図 3 - 6 に

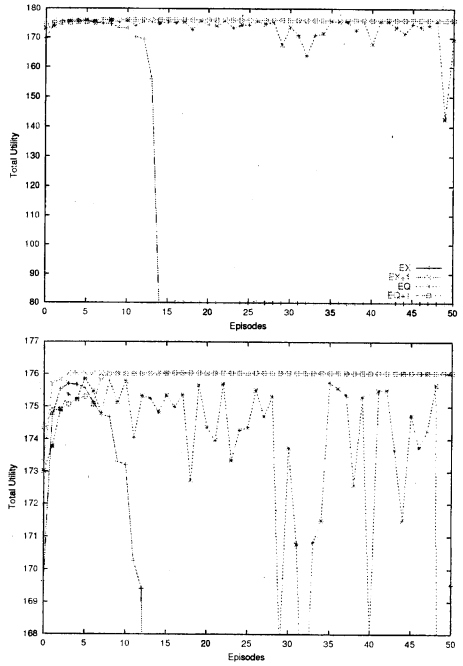

図 2 全体の効用合計の変化 (下は拡大図)

示す。EX では学習は全くうまく行かないのに対し、EX+1 では短期間の間に Nash 均衡へ収束することに成功した。エージェントの分布を見ると、EX では非常に不安定な状態が続き、14 エピソード目において、1 つの資源に全エージェントが集中してしまい、その状態から抜け出せなくなっていることが分かる。これは、全エージェントが同じ資源を選択してしまった結果、他の資源に対する情報が全く入って来ない状態に陥ったことを意味する。それに対し、EX+1 では、選択人数は Nash 均衡の分布に向かって大きな順位の変動もなく収束していることが見てとれる。一方、EQ については EX と比べると穏やかではあるが、いつまでも学習が安定しない状態が続き、全体の効用合計は平均的に低い値を示している。これに対し、EQ+1 は、途中、多少の振動はあるにせよ、最終的にはほぼ EX+1 と同様の結果に収束し、予想通りの結果を示した。

5. 考　　察

実験により、大域情報の悪影響を確認し、提案手法の有効性を確認することができた。また、EQ+1 に関する考察も予想通りの結果となることを確認した。提案手法は、全体の人数や Nash 均衡時の人数配分が分

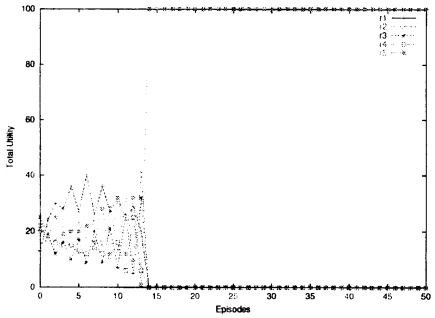

図 3　各資源の選択人数の変動 (EX)

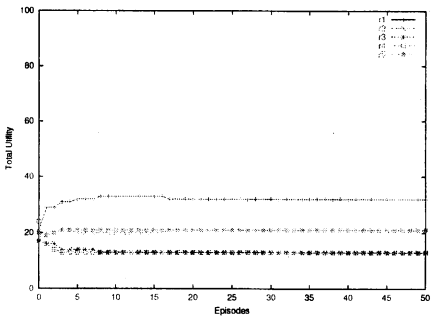

図 4　各資源の選択人数の変動 (EX+1)

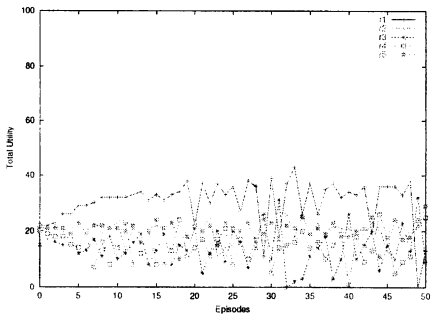

図 5　各資源の選択人数の変動 (EQ)

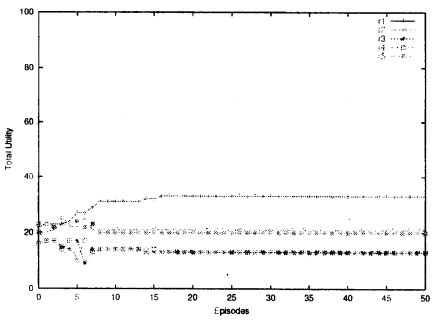

図 6　各資源の選択人数の変動 (EQ+1)

かっていなくても、個々の資源を選択した人数さえ分かれば、大域情報の悪影響を悪影響を避けつつ、Nash 均衡に収束するという利点がある。この性質は、現実世界で用いる場合には、非常に有利に働くものと思われる。

提案手法は、Nash 均衡に対する収束性と安定性は保証しているが、最適解への収束は必ずしも保証していない。そのため、最適解を求めるためには、更に別の手法が必要となる。しかし少なくとも局所的には最適であるということは充分に有益なことであり、また、エージェントの選択が安定しているということは行動の変更にコストがかかるような環境においては非常に重要である。

この手法は、現実的には、複数の Web Server における負荷分散や、複数の経路が存在するナビゲーションにおける混雑状況の調整などに用いることが可能である。この場合、提案手法はユーザに多少加工した情報を流すことにより調整を行なうことに相当する。その情報を見て、人間が実際にどのように行動を変化させるかということに関しては本稿では一切議論をしていないが、エージェントが全て利己的に行動するという前提があれば、同様の問題が生じ、それに対して提案手法に類似した方法が有効となる可能性は高いものと思われる。

本研究では、問題を最も単純化した場合の実験についてのみ示した。これを元に、以下のような設定を追加して行くことが将来課題として考えられる。

- エージェントが探査的行動を行ない、自分でより良い選択肢を探す要素が含まれる場合
- 大域情報の悪影響が既に存在している環境
- エージェントが個々の資源に対して、好みの要素を入れた場合

6. おわりに

利己的なエージェントによるマルチエージェント環境において、大域情報の悪影響を軽減する手法を提案した。この手法は、個々のエージェントレベルで問題を解決するのではなく、エージェント間で流れる情報を加工することにより全体を制御するという点に特徴がある。提案手法は、現状をそのまま伝えるのではなく、現在の分布に対し 1 人多い場合の効用を大域情報として伝えるという単純な方法であり、これには、全体の人数や Nash 均衡時の人数配分など、問題固有の情報が分からなくても適用できるという利点がある。

本稿では、この方法の Nash 均衡での安定性、および Nash 均衡への収束性を示し、資源配分問題を用いた実験において、その有効性を確認した。実験には非常に単純な問題を用いたが、この手法は実世界への適用が比較的容易であり、将来的には、より現実的な環境への適用を目指している。

参 考 文 献

1) S. Bankes. Exploring the foundations of artificial societies: Experiments in evolving solutions to iterated n-player prisoner's dilemma. In *Artificial Life IV*, pages 337–342. 1994.
2) D.Challet and Y.-C. Zhang. Emergence of cooperation and organization in an evolutionary game. *Physica A*, 246(3-4):407–418, 1997.
3) K. Izumi and K. Ueda. Phase transition in a foreign exchange market-analysis based on an artificial market approach. *IEEE Transactions on Evolutionary Computation*, 5(5):456–470, 2001.
4) F. Klügl, A. L. C. Bazzan, and J. Wahle. Selection of information types based on personal utility - a testbed for traffic information markets. In *Proceedings of the Second International Joint Conference on Autonomous Agents and Milti Agent Systems*, pages 377–384, 2003.
5) J. Schneider, W.-K. Wong, A. Moore, and M. Riedmiller. Distributed value functions. In *Proceedings of the Sixteenth International Conference on Machine Learning*, pages 371–378, 1999.
6) M. Tan. Multi-agent reinforcement learning: Independent vs. cooperative agents. In *Proceedings of the Tenth International Conference on Machine Learning*, pages 330–337, 1993.
7) T. Yamashita, K. Izumi, K. Kurumatani, and H. Nakashima. Smooth traffic flow with a cooperative car navigation system. In *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Milti Agent Systems*, pages 478–485, 2005.
8) 栗原聡. Minority game の不思議. 情報処理学会誌, Vol.45(No.4):388–394, 2004.