原著論文

異文化体験ゲームにおける 集団適応エージェントの開発†

大村 英史 *1・片上 大輔 *1・新田 克己 *1・野澤 孝之 *2・近藤 敏之 *2

エージェントの人間の社会への進出に伴い、人間とエージェントの間の相互作用(HAI: Human-Agent Interaction)の研究が活性化している。しかしながら、社会の中での人間とエージェントの関わり方の方法論は確立されていない。そこで、本論文では社会適応の一つの手段として「集団内の暗黙のルール獲得」に注目し、社会的な集団適応エージェントを実現することを目的としている。社会集団にはルールが存在していると言われている。もし集団に属するあるメンバーがこのルールに従わない場合、集団の他のメンバーと協調できないばかりか攻撃されかねない。このように、集団に適応するためには、その集団におけるルールを獲得しそれに従うことが重要であると言える。そこで、人間の集団内のルール獲得を観察し、その知見を用いて集団内のルール獲得によって集団適応を行うエージェントの設計を提案する。

はじめに、集団ルール獲得ができる環境を構築する。これは、異文化体験シミュレーションゲーム *BARNGA* をベースに *Online BARNGA* として作成する。この環境内で人間同士によるゲームを行い、人間のルール獲得の過程の観察を行う。これにより、人間が暗黙の集団ルールを獲得する際には、「気づきの有無」と「ルールに従った振舞いの有無」により分類できる3つの内部状態間の遷移を行うことと、暗黙の集団ルール獲得には、勝者判定規則の獲得とゲームに勝つための集団戦略獲得が存在することの2つの獲得があることが観察できた。次に、人間の暗黙のルール獲得に関わる内部状態の遷移のタイミングを調べるために、fNIRS(機能的近赤外線分光法)を用いて、実験時の被験者の前頭葉の脳活動計測を行い、状態遷移がどのタイミングで行われたかの知見を得た。これらの知見を用いて、「気づき」と「振舞い」による3つの内部状態の遷移と、勝者判定規則と集団戦略を別々に学習するエージェントの設計方法を提案する。さらに、このエージェント同士で *Online BARNGA* 上でゲームシミュレーションを行い、人間の知識獲得のプロセスの結果と比較を行う。これにより、提案したエージェントが人間と同じ戦略に収束していることから、人間と同じ暗黙のルールへの適応を行う機能を持っていることを示すことができ、これをもって集団適応エージェントの提案とする。

キーワード：HAI, 社会適応エージェント, 集団適応

1. はじめに

近年、人間の生活をより快適にするために、さまざまな面でエージェントが人間の生活にかかわりつつある。そのため、人間とエージェントの関わり方に焦点を当てたヒューマン・エージェント・インタラクション(HAI) [1]と呼ばれている研究領域は活発であり様々な研究がおこなわれている。HAIでは1対1のインタラクションとして考慮することが中心であったが、人間の活動する場を踏まえた社会的な活動について議論されるようになってきている [2, 3]。しかしな

がら、社会における人間とエージェントの関わり方の方法論は確立されていない。そこで、我々は、人間社会は複数人で構成されているため、多人数とインタラクションを行うことを考慮する必要があることに着目し、多人数いる中で1対1のインタラクションを通して効率よく複数の相手に適応するエージェントの開発を行ってきた [4, 5]。これにより、社会を構成するメンバーの情報を利用し、個人適応を効率よく行うことを実現することができた。しかしながら、このエージェントはあくまで個人適応であって、社会全体の考慮はされていない。社会適応のためには社会全体を考慮する必要がある。

一般に、社会のような集団には集団全体に共通のルールが存在し、集団に適応するためには、集団内に存在するルールを獲得しそれに従っていかなくてはならない。もし集団に属するあるメンバーがこのルールに従わない場合、集団の他のメンバーと協調できないばかりか攻撃されかねない。そこで、我々は、HAI研究の一環として、エージェントが人間の社会へ適応で

† Development of Social Adaptive Agents in Simulation Game of Cross-Cultural Experience

Hidefumi OHMURA, Daisuke KATAGAMI, Katsumi NITTA, Takayuki NOZAWA and Toshiyuki KONDO

*1 東京工業大学 大学院総合理工学研究科

Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology

*2 東京農工大学 大学院共生科学技術研究院

Institute of Symbiotic Science and Technology, Tokyo University of Agriculture and Technology

きるエージェントを作成するために、人間が集団へ適応するときに利用する社会的スキルについて分析を行ってきた[6, 7]。我々は、集団内に存在するルールのなかの一つである暗黙のルールの獲得に関して調査を行った結果、人間は「気づき」と「振舞い」を適切に行うことで暗黙のルールの獲得をしていることが分かってきた。また、暗黙のルールには規則の変化と道徳の2種類あることも確認できた。この知見を用いて、我々は社会のような集団適応のために2種類の暗黙のルールを獲得できるようなモジュール設計と、「気づき」と「振舞い」による3つの状態に遷移を実装したエージェントの設計を行う。このエージェントは、社会適応が必要な異文化体験シミュレーションゲーム *Online BARNGA* 上で活動できるように設計を行う。本論文では、人間の集団適応の観察から得られた知見を用いて実現する集団適応エージェントの紹介を行う。そして集団適応エージェントと人間の行動の比較した結果を示しエージェントの集団適応を示す。

第2章では、社会適応エージェントに実装する社会へ適応するための能力を説明する。また、人間のその能力が観察できるシミュレーションゲーム *Online BARNGA* を紹介する。第3章では、人間の被験者を用いた *Online BARNGA* での実験を示す。第4章では、その知見を用いた集団適応エージェントの設計について説明し、エージェントによる *Online BARNGA* の実験の結果を示す。第5章では、人間による実験とエージェントによる実験の比較を行い、集団適応エージェントにおいて考察を行う。最後に第6章でまとめる。

2. 社会的能力

2.1 社会と社会的能力

社会の定義の一つに、合意に基づき社会的協力の理念と共通のルールを共有する人々からなる集団、というものがある[8]。藤川によると、人間たちは、なにも存在しない自然状態から、共に理念や共通のルールを規制として社会を形成していく。自然状態では外部環境によって様相は変化するが、外部状態が厳しいと、悪徳が増大し戦争状態に陥る。戦争状態を回避するため社会的協力の理念と共通のルールの規制を導入する。また、全体主義を追求すると規制が大きくなり、個人の自由が小さくなる。個人が自由を得るために共通のルールの崩壊が起こる。つまり社会の規制は個人の利得の合計が最大になるような均衡が理想である。山本は、戦時中の日本での全体主義の恐ろしさについても紹介している[9]。ここで我々が注目することは、社会には理念やルールという規制が存在し、人間は常にそれを取り入れようとしていることである。

また、人間の社会と接するための能力を測る基準として、いくつか考案されているものがある。例えば、ピーター・サロヴェイらは、EQ (Emotional Intelligence Quotient) [10] という概念を理論化している。EQは人間の知性の一つであり大きく二つに分かれている。一つ目は、自分の感情を知り、現実的な自己モデルを形成してそれを行動指針とする能力(心内知性)である。二つ目は、周囲の人の同調し、欲求を捉えて適切な行動をする能力(対人知性)である。また、近年、精神科の病院でSST(social skills training) [11] と呼ばれるトレーニング方法が行われている。これは認知行動療法の一つであり、社会的能力のトレーニングである。社会的能力とは、状況に合わせて、言語的・非言語的行動を上手に用いて、自分の気持ちや考えをうまく人に伝える能力のことである。EQの対人知性とSSTの社会的スキルは、両方とも利害衝突の回避のための共通のルールを獲得する能力とも考えることができる。我々はこの共通のルールの獲得に着目し、この能力のことを社会的能力の一つであると考え、

脳科学の分野でも、人間の社会的な活動について注目されており、このような活動するための脳活動が確認されている[12]。特に「扁桃体」や「腹内側前頭皮質」などが、脳の中で社会的能力にかかわっているとされている。

2.2 社会におけるルール

人間が集団でいるとき、不安定な弱肉強食な状態から自らの利益を得るためにルールを設けるといわれている[8]。このとき発生したルールを獲得することができないだけでなく、利益を得る方法もわからなければ集団の安定状態を壊してしまい他のメンバーから攻撃されかねない。そのため、集団に適応するために、集団に存在するルールを獲得しなくてはならない。

Serifらは心理実験において、人間は集団内にいると集団の圧力によりルール(社会規範)を獲得することを示している[13]。エージェントにおいても、報酬に基づいた強化学習により、利害の衝突回避のため社会性が発現することも確認されている[14]。衝突回避のため社会性は集団に存在するルールを獲得しているといえるであろう。

ここで、社会におけるルール(社会的協力の理念と共通のルール)を規範に照らし合わせて規則と道徳の二つに分けてまとめる(表1)。規則が共通のルール、道徳が社会的協力の理念に相当する。まず、規則としての規範について説明する。規則の例として法律や命令などがあげられる。規則は制度化されているため、一意に決まっており、内容も公表されていることがほ

表1 ルールの分類

規範	規則	道徳
例	法律, 命令 約束, 契約	マナー, 常識 不文律, 場の空気
特徴	制度化されたルール 一意に決まっている	制度化されていない 一意に決まっている
獲得法	受動的	能動的

とんどである。そのため、この内容は直接教示として得ることができるので、受動的に獲得ができる。次に道徳としての規範について説明する。道徳の例はマナーや場の空気などがある。道徳は、制度化されていないため、集団毎に異なっていることが多く、公表されていないことがほとんどである。そのため、自らの経験を通してルールを能動的に見つけなくてはならない。道徳としての規範獲得は規則としての規範獲得に比べて難しいと考えられる。また、規則の中でも、国などの社会集団を超えるときや、時代による解釈の違いにより道徳と同様に変化が生じるため、能動的に獲得しなくてはならないこともある。そこで我々は、能動的な獲得を必要とする規範を暗黙のルールと名付け、このルールの獲得を社会的能力の一つとして着目する。

2.3 社会適応ゲーム

我々は、社会的能力の一つである暗黙のルールの獲得する能力に着目し観察を行った。この暗黙のルールを獲得することを観察するために、異文化体験シミュレーションゲームBARNGAを基にOnline BARNGAを作成した。まず、BARNGAで用いられるトリック・テイキング・ゲームの説明を行い、次にBARNGAの説明を行い、最後にOnline BARNGAの説明を行う。

2.3.1 トリック・テイキング・ゲーム

トリック・テイキング・ゲームはトランプゲームの一つである。一般的にトランプゲームでは、スペード、ハート、クラブ、ダイヤのことをスーツ、数字やA, J, Q, Kのことをランクという。各プレイヤーに同じ枚数のカードが配られる。そして、手札がなくなるまでトリックとテイキングを繰り返しゲームは進行していく。トリックでは、定められた親から順に一人ずつ場に出していく。全員がカードを出し終わったら、場に最も強いカードを出したプレイヤーがトリックの勝者となり、場にあるカードを持っていく。これがテイキングである。トリックとテイキングを繰り返し行い、獲得したカードが最も多いプレイヤーがトリック・テイキング・ゲームの勝者となる。

トリック・テイキング・ゲームのカードの強さは

ゲームによって異なっている。ゲームによっては切り札というものが設定されている。切り札とは定められたスーツのことであり、このスーツはほかのスーツよりも強い。

2.3.2 BARNGA

BARNGA[15]はシミュレーションゲーム[16]であり、文化的差異を仮想的に体験することを目的としている。このゲームは、トランプのルールを文化と見立て、その違いに気がついたときの驚きをカルチャーショックとして体験する、としている。

このトランプゲームのルールは一般的な社会における規範と比べると規模が小さく単純であり、社会の再現としては不十分である。しかし、本研究で注目している規則の変化の暗黙のルールが含まれていることから、プレイヤーの暗黙のルールの獲得が観察できるのでこのゲームを採用した。

ゲームの内容を説明する。ゲーム中はプレイヤー同士会話してはならない。このため、プレイヤー達は相手に自分の思っていることを、ジェスチャーなどのノンバーバル情報を用いて伝えなくてはならない。まず、テーブルが複数存在し、各テーブルに4、5人のプレイヤーが着席する。全員が席に着いたらトランプゲームのルールが書かれた紙が配布される。ここで伝えられるトランプゲームのルールがテーブルごとに微妙に異なっていることである。異なっている点はカードの強さやゲームのターンの回し方など単純なものである。ただし、プレイヤーには各テーブルでトランプゲームのルールが異なっていることは伝えられていない。全員が読み終わると、トランプゲームが開始される前にトランブルールの書かれた紙が回収される。そのため、ゲーム中はルールを参照することはできない。同じテーブルに、着席しているプレイヤー同士でトランプゲームを行う。ある程度トランプゲームを行ったら、各テーブルで、トランプゲームを中断して、テーブル内で最も勝っているプレイヤーと負けているプレイヤーがほかのテーブルのプレイヤーとトレードされる(図1)。トレード終了後、各テーブルでトランプゲームを再開する。このとき、トレードされたプレイヤーは、異動後のテーブルにて移動前のテーブルと異なったルールでゲームが進行しているため、違和感を覚える。また、トレードされていないプレイヤーも、トレードされたプレイヤーの行動に対して違和感を覚える。BARNGAでは、一切の会話が禁じられているため、この違和感を解消する方法として、周りの動きを観察したり、ジェスチャーなどのノンバーバル情報を用いてトランプゲームを進めていかなくて

図1 トレードの方法

はならない。この違和感をBARNGAでは異文化体験と見立てている。

我々の定義した暗黙のルールはBARNGAの中ではルールの違いとして考えることができる。そのため、このゲーム中のプレイヤーを観察することにより暗黙のルールの獲得を観測できる。BARNGAは集団適応のためのエージェント研究としても用いられている[17]。しかしながら、このゲームでのプレイヤー同士のインタラクションはあいまいなものであり、ゲーム中のプレイヤーの行動ログデータの確保は難しい。そこで、我々はプレイヤーのログデータの記録とインタラクションを明確化のためにOnline BARNGAを作成した。

2.3.3 Online BARNGA

Online BARNGAは、BARNGAをオンライン化により、プレイヤーの行動ログデータの記録とプレイヤー同士の明確なインタラクションのための勝者の決定方法と得点制を導入したものである。この得点の調節により、規則としての暗黙のルールだけでなく道徳としての暗黙のルールが発生させることができる。

まず、Online BARNGAでは、オンライン化により各プレイヤーはクライアントPCからサーバPCに接続しゲームを行う。クライアントPCは図2のようなインターフェースになっており、各プレイヤーはこのインターフェースを操作してゲームを進める。サーバPCにて、ゲーム中の各プレイヤー状態(手札、場札、ランキングなど)や行為(出したカード、勝者の選び方など)をログデータとして記録する。

また、明確な勝者の決定方法の導入により、社会的な振舞いをデータとして取得できる。これを実現するために、得点の導入をしている。勝者決定にかかわる手続きと、得点の変動の流れを以下に示す。

図2 クライアントPCのインターフェース

1. テーブルから一人親が選ばれる
2. 親から順にプレイヤー全員カードを場に出す
3. 親が勝者を決定する
4. 勝者にWinner Pointが入る
5. 子は親の決定に従うかクレームをつけるか決定する
6. クレームをつけた子からクレーム賃としてExpense Pointを減点する
7. 親と勝者からクレームのついた割合を重みとしたPenalty Pointを減点する
8. クレームをつけた人はクレームのついた割合を重みとしたBonus Pointを加点する

また、Online BARNGAでは、各得点の値を変えることにより、ゲーム設計者が故意的に偏った状況(戦略的にルールに従わないなど)も作り出すことができる。実験ではこの得点を2通りに設定した。

3. 人間の集団適応

3.1 人間同士のOnline BARNGAの実験

3.1.1 実験設定

我々は、人間の暗黙のルールの獲得を観察するために、Online BARNGAを用いた実験を行った[6]。15人の被験者で3テーブルでOnline BARNGAを行った。実験状態は、得点とトレードまでのカードを配る回数の設定により2種類の状態を設定した。1つ目の設定は秩序状態と名付け、テーブル移動後のプレイヤーもトランプゲームのルールに従うように、次の条件を満たすように得点が設定してある。一人しかクレームをつけなかった時には大きく損をする。クレームがない時よりある時のほうが全員の総得点が小さい。3人以上クレームをつけた時は親よりクレーマーのほうが得をする。自分を勝者にしてもクレームがつ

表2 得点変動値

ポイント名	説明	設定 1	設定 2
Winner Points (WP)	勝者が得るポイント	400	100
Penalty Points for Dealer ($r \cdot PP_D$)	勝者に対するペナルティポイント	400	80
Penalty Points for Winner ($r \cdot PP_W$)	親に対するペナルティポイント	80	80
Bonus Points ($r \cdot BP$)	クレイマーが得るポイント	200	40
Expense Points (EP)	クレームをつけるためのポイント	40	20

表3 トランプゲームのルールの違い

	切り札	ランクの強さ
Table1	ハート	$7 > 6 > 5 > 4 > 3 > 2 > A$
Table2	スペード	$A > 7 > 6 > 5 > 4 > 3 > 2$
Table3	クラブ	$2 > A > 7 > 6 > 5 > 4 > 3$

けば損をする。2つ目の設定は混乱状態と名付け、各プレイヤーがトランプゲームのルールには従わず戦略的に振る舞えるよう、全員クレームがついたときのみ親は損をするという条件で得点を設定した。各得点は表2に示す。また、設定1の秩序状態ではトレードまでにカードが配られる回数を4回に設定しており、設定2の混乱状態では1回に設定してある。

また、各テーブルにて初期に配られるトランプゲームのルールはカードの強さが異なっている。各テーブルにおける切り札とランクの強さを表3に示す。

3.1.2 実験結果

実験の結果、2つの状態で異なる暗黙のルールが観察された。設定1（秩序状態）ではテーブル移動したプレイヤーにとって、移動後のテーブルのトランプゲームのルールが暗黙のルールとなった。一方、設定2（混乱状態）では、トランプゲームのルールは暗黙のルールではなく、戦略的な行動が暗黙のルールとなった。その戦略としての暗黙のルールは「親のときは自分を勝者にする」と「共にクレームをつけてボーナスポイントを得る」というものである。この2種類の暗黙のルールを規範に照らし合わせると、トランプゲームのルールの変化は規則の変化であり、戦略的な行動の暗黙のルールは道徳として考えることができる。

また、人間は暗黙のルールを獲得する際、「気づき」と「振舞い」を行っている。この二つ行為により3つの状態を作っている。この3つの状態を表4にまとめる。状態3は、暗黙のルールが存在することに「気づかない」状態である。状態2は、暗黙のルールが存在することに「気づいて」いるが、暗黙のルールの内容を理解していないためルールにどおりに「振舞わない」状

表4 振舞いと気づきによる状態

	←社会的		非社会的→	
	状態 1	状態 2	状態 3	
気づき	暗黙のルールの存在に気づく		気づかない	
振舞い	暗黙のルールに従う		従わない	

態である。状態1は、暗黙のルールが存在することに「気づいて」おり、暗黙のルールの内容を理解しルール通りに「振舞う」状態である。状態3は、社会集団の変化に気づかず、集団とは関係のない基準で行動をしているため、非社会的といえる。一方、状態1は集団内の暗黙のルールを発見し獲得を行い、集団内の基準に従って行動しているため社会的といえる。

3.2 人間の暗黙のルールの獲得時の脳活動計測

3.2.1 実験設定

人間の社会的スキルの一つである暗黙のルールの獲得は「気づき」と「振舞い」によって生じる。この社会的スキルを伴った行動を起こしているときの脳活動を計測し、どのタイミングで「気づき」が起きているのかを確認するのが本実験の目的である。

実験は、3人の被験者が順に行う。実験の環境は、被験者1人とエージェントプレイヤー14体、3つのテーブルでOnline BARNGAを行う。エージェントプレイヤーは最初に与えられるトランプゲームのルールに従ってゲームを進めるが1/10の確率でランダムな行動をするように設定してある。配られる手札とエージェントが場に出すカードは、事前に行った人間同士の実験で得られたデータをもとに作成した。これらの設定は、被験者が人間と対戦しているように感じさせるためと、被験者間の計測データを同じ環境で比較するためである。また、この実験でのOnline BARNGAのトレードは各テーブルから1度に1人だけであり、被験者が選ばれるようになっている。このとき、被験者にはトレードされるプレイヤーはランダムで選ばれらる伝えてあり、被験者は自分がトレードされることを知らない。被験者の学習が十分できるように、トレードまでのトランプが配られる回数を6回に設定した。実験は実験時間が1時間に達した時点で終了する。実験後にOnline BARNGAについての感想をアンケートにより答えてもらう。

3.2.2 脳活動計測：fNIRS

脳活動計測の手法として、fNIRS(機能的近赤外線分光法)を用いた。これは、頭部に近赤外線を照射して反射率を計測することで血中の酸素化ヘモグロビンの変化を計測する手法である。脳血流の増加はその部

図3 脳活動計測の様子

図4 測定部位

位の神経活動の増加を反映している[18]。測定器は島津製作所FOIRE-3000を用いる。測定部位は社会的な意思決定をしているといわれる前頭葉[12]を図るために、鼻と後頭結節および左右の耳の間に均等に分割し頭部計測の部位を定義する国際10-20法のFz(前頭中央の位置)を頂点となるよう22チャンネルで行った(図4)。

3.2.3 実験結果

3人被験者はそれぞれA, B, Cとする。また、トレードまでのトランプゲームの期間をステージと呼ぶ(各ステージで6回トランプが配られる)。被験者A, Bは2回目のトレード直前で実験時間が1時間となり2ステージで実験を終了した。被験者Cは3ステージ終了まで実験を行った。アンケートによると被験者Aはテーブル間でトランプゲームのルールが異なっていることに気づくことができなかったが、被験者B, Cは気づくことができた。ルールの変化に気づかなかったAの脳計測の結果は目立った変化は見られなかった。ルールの変化に気づいた被験者Bの2ステージと、被験者Cの2, 3ステージに着目する。それぞれ親のときに勝者の決定を行ったときに注目し、勝者の決定をしてから30秒間の酸化ヘモグロビン(oxy-Hb)の変化を示す(図5, 6, 7)。0秒で勝者を選択しこのときをoxy-Hbの値を0にセットしてあ

図5 被験者Bのステージ2における、勝者を選択したときの脳活動の変化(19チャンネル)

図6 被験者Cのステージ2における、勝者を選択したときの脳活動の変化(19チャンネル)

図7 被験者Cのステージ3における、勝者を選択したときの脳活動の変化(19チャンネル)

る。8秒から10秒後に他のプレイヤーからの評価(クレーム)を受けている。それぞれの結果は左右下部の19, 14, 18, 22チャンネルの部位が活発に活動しており、その中でももっとも特徴の出ている19チャンネルの値である。

ここでは、その時に着いているテーブルの初期に与えられているトランプゲームのルールに従わずに勝者決定をすることを、「間違えている」と表記する。被験者Bの2ステージでは、1回目、2回目、5回目、6回目にてクレームが多く発生している。1回目のときは間違えているが脳活動の変化は見られない。そのため、このとき気づきは生じていないと考えられる。2回目のときも間違えている。脳活動は評価の後に一度下がりその後上昇している。この活動が気づきの活動と考えられる。5回目は自分の被験者の選択ミスであった。クレームの多さに対して脳活動が下がっているが、自分のミスによるものと理解しているため、脳活動が上がらないと考えられる。6回目では間違いをしていない。しかしながら、エージェントのランダム行動によりクレームが付いている。自分の選択でクレームは付かないと確信している時に、クレームをつけられたことに大きなショックを受けていると考えられる。そのため一度下がりその後上昇したと考えられる。

被験者Cの2ステージでも2回目と、3ステージの1回目において上記と同じような現象が見られた。これらはテーブルのルールが異なっていることに対する気づきの活動であると考えられる。また、5回目と6回目においても同様な活動が見られる。被験者の戦略的な行動に対してのクレームである。自分の戦略が有効ではないことへの気づきであると考えられる。

この結果より、人間は、他のプレイヤーから評価(クレーム)によって脳活動が起こっており、「気づき」が生じていると考えられる。エージェントの設計には、評価を報酬として気付きによって学習を生じさせればよいことを示唆している。

4. 集団適応エージェント

4.1 概要

人間の集団適応の実験で得られた知見を用いてエージェントの設計を行う。まず、暗黙のルールは、規則の変化としての暗黙のルールと、戦略としての暗黙のルールが存在するため、両方を獲得できるように設計を行う。そして、それぞれの暗黙のルールの獲得のために「気づき」と「振舞い」を用いて3つの状態を遷移できるように設計を行う。遷移は他のプレイヤーからの評価を報酬とした学習で行う。

図8 エージェントの各モジュール

4.2 Online BARNGAで活動するための4つのモジュール

集団適応エージェントはOnline BARNGA上で活動できるように設計をする。Online BARNGAでは、各プレイヤーは3つの行動を行う。それぞれの行動における戦略の学習が行えるように、各行動に対してモジュールの設計を行う。またそれぞれの行動は、トランプゲームのルールに依存する場合がある(しない場合もある)。そのため、トランプゲームのルールを学習するモジュールを各行動モジュールの上位に配置し、エージェントが獲得したトランプゲームのルールが下位のモジュールに流れるように設計をする。設計の概要図を図8に示す。各モジュールについて説明をする。

4.2.1 トランプルール学習モジュール

エージェントは、このモジュールでトランプゲームのルールを学習する。学習は、ランク学習とスーツ(切り札)学習を独立して行う。学習システムは、強化学習(Q学習)[19]を用いる。ランク学習では、AからKまでの13種類において評価値を用意する。スーツ学習では、スペード、ハート、クラブ、ダイヤの4種類において評価値を用意する。評価値の最も大きいものが最も強いカードとなる。評価値の更新は、加点された時は1、減点された時0とした報酬で行うものとする。また、ランク学習は同じスーツと比較していないと意味がないので、勝ったカードと同じスーツが場に出ている割合を重みとして学習率にかける。スーツの学習における学習率は0.4、ランクの学習における学習率は0.6に設定をした。

4.2.2 手札選択戦略モジュール

エージェントは、このモジュールで手札から場に出す捨て札の選択を行う。Online BARNGAのトランプゲームにおいて、手札の選択は重要な戦略となる。しかし、人間同士の実験では、手札選択による戦略から

暗黙のルールは発生しなかった。そのため、戦略の学習は行わずに、基本的な戦略にのっとって行動するように設計を行った。基本的な戦略の例は、強いカードは自分のターンが最後の時まで温存する、ターンが早い時は弱いカードを出す、などである。

4.2.3 勝ちカード選択戦略学習モジュール

エージェントは、親のときにこのモジュールで勝ちカードを選択する。勝ちカードの選択は、人間同士の実験でトランプゲームのルールに従い選択する方法と、戦略的にトランプゲームのルールに依存せず選択する方法が確認された。そのため、トランプゲームのルールに従う戦略と、トランプゲームのルールに従わずに勝ちカードを選択する戦略の両方が取れるような設計を行った。学習システムにはXCS(クラシファイアシステム)[20]を用いる。これは生成されたルールを認識しやすく、他のエージェントと簡単に比較できるからである。また、XCSの“don't care”により入力情報の無視が簡単に実現できるからである。入力情報は、自分のターン(何番目か)、各プレイヤーの順位、各プレイヤーのカードを出す順番、場札である。出力情報は、トランプゲームのルールに従う、またはどのカードを勝ちにするか、である。トランプゲームのルールに従う、または、トランプゲームのルールを無視して場に出ている任意のカードを勝ちにするかである。トランプゲームのルールに従うが出力された場合は、トランブルール生成部で生成されたトランプゲームのルールを用いて勝ちカードを選択する。学習は「勝ちカード決定に対するクレームの数」が入力されたときこの値を報酬として行う。また、学習率は0.3に設定した。

4.2.4 クレーム戦略学習モジュール

エージェントは、子のときにこのモジュールで親の決定に対してクレームをつけるか否かを選択する。このモジュールも、勝ちカード選択部と同様にトランプゲームのルールに従い選択する方法と、戦略的にトランプゲームのルールに依存せず選択する方法が人間同士の実験で確認された。そのため、トランプゲームのルールに依存する選択と、トランプゲームのルールを無視して戦略的な選択と両方が取れるように設計を行っている。学習システムも勝ちカード選択部と同様にXCS(クラシファイアシステム)を用いる。入力情報は、各プレイヤーの順位、各プレイヤーのカードを出す順番、場札である。出力情報は、トランプゲームのルールに従う、またはクレームをつけるか否かである。トランプゲームのルールに従う場合は、親の選択

した勝ちカードと、このエージェントのトランブルール生成部で生成されたルールを用いて選択した勝ちカードが異なるときにクレームをつける。学習は「クレームに対する得点」が入力されたときこの値を報酬として行う。また、学習率は0.3に設定した。

4.3 エージェントの状態の設計

人間同士のOnline BARNGAにおいて、人間は暗黙のルールを獲得する際、「気づき」と「振舞い」により3つの状態をとることを確認した(表4)。状態1は、暗黙のルールを獲得し、暗黙のルールでゲームをおこなっている状態である。状態2は、暗黙のルールを学習している状態である。状態3は、暗黙ルールを知らずにゲームを行っている状態である。ここで状態1と状態3を比較してみると、状態1において保持しているルールは他のエージェントと同じルールであり、状態3において保持しているルールは他のエージェントと異なるルールである。つまり、状態1と3はエージェントの状態としては同じであり、保持しているルールが異なるだけである。そのため、状態1と状態3は学習終了状態としての設計をし、状態2では学習状態として設計を行う。状態1および状態3では、この状態に遷移したときに保持している学習結果を基に行動を決定する。状態2ではルールセットおよび変数は常に更新をし、更新後の値で行動を決定する。

また、状態の遷移を、状態3から状態2(遷移A)と、状態2から状態1(遷移B)と、状態1から状態3(遷移C)とする。遷移Aおよび遷移Bは学習に依存されるため、閾値を設け、この値を超えた時に遷移する。遷移Cは周りのプレイヤーが保持するルールとの比較で遷移するので、自動的に遷移される。これらの状態遷移を図9に示す。設定された閾値の大きさと、遷移のされやすさが決まるため、人間の実験で得られた個人差が表現できる。

図9 エージェントの状態遷移

5. 暗黙のルールを獲得するエージェントの評価実験

5.1 実験設定

本実験では、設計を行った暗黙のルールを獲得するエージェントがOnline BARNGAのなかで暗黙のルールを獲得する様子を観察する。その様子を人間のルール獲得実験と比較を行い、その有用性について評価を行う。

実験条件は、テーブル移動までのカード配布回数を3種類と、得点設定を2種類の組み合わせで行う。テーブル移動までの配布回数は4回(条件A)、2回(条件B)、1回(条件C)とする。得点設定は表2の設定1と設定2と同じ条件を、それぞれ条件aと条件bとする。各条件の組み合わせで条件を表記する。人間同士のOnline BARNGAの実験の設定は、設定1(秩序状態)の時は条件Aaとなり、設定2(混乱状態)の時は条件Cbとなる。

5.2 実験結果

5.2.1 秩序状態における実験結果

秩序状態におけるエージェントの動作結果の例を示す。トランブルール学習モジュールのスーツのQ値の変化を図10に示す。

トレード後にテーブルのトランプゲームのルールであるQ値の値が大きくなっており、学習出来ていることが分かる。ランクの結果も同様なものとなり、学習できていることが分かった。

勝ちカード戦略モジュールの各戦略における予測値の平均値の変化の例を図11に示す。トランプゲームのルールに従うという戦略の予測値が大きくなっており、人間の結果と同じ戦略に収束していることが分かる。

また、各戦略モジュールでも人間同士の実験の結果と同様なものとなった。

図10 秩序状態におけるスーツのQ 値の変化

図11 秩序状態における勝ちカード戦略の予測値の変化

5.2.2 混乱状態における実験結果

混乱状態におけるエージェントの動作結果の例を示す。トランブルール学習モジュールのスーツのQ値の変化の例を図12に示す。

スーツのQ値もランクのQ値もトレード後にテーブルの初期に与えられたトランプゲームのルールに依存せず、ばらばらな値をとっている。これより移動後のテーブルのトランプゲームのルールを学習出来なかったことが分かる。

勝ちカード戦略モジュールの各戦略における予測値の平均値の変化の例を図13に示す。特に目立った戦略を得ることができてはいないが、人間の実験で出た自分を勝者にするという戦略はここでは1枚目につけるということであり、この予測値がわずかではあるが大きくなっていることはこの戦略が暗黙のルールになりつつあるといえる。

クレーム戦略モジュールの各戦略における予測値の平均値の変化を図14に示す。クレームをつけるという戦略の予測値が大きくなっている。この戦略が暗黙のルールとして収束していることが分かる。

人間同士の実験とエージェント同士の実験の集団安定の戦略を表5に示す。過半数を超えたものを暗黙の

図12 混乱状態におけるスーツのQ 値の変化

図13 混乱状態における勝ちカード戦略の予測値の変化

図14 混乱状態におけるクレーム戦略の予測値の変化

表5 暗黙のルールとしての戦略の比較

	秩序状態		混乱状態	
	エージェント	人間	エージェント	人間
トランプルール学習	トランプルール学習	トランプルール学習	トランプルール学習できない	トランプルール学習できない
勝ちカード選択戦略学習	トランプルールに従う	トランプルールに従う	戦略なし	自分を親にする
クレーム戦略学習	クレームを付けない	トランプルールに従う	クレームをつける	クレームをつける

ルールとしている。表からわかるように、秩序状態のクレーム戦略学習と、混乱状態の勝ちカード選択学習が異なる暗黙のルールとしての戦略となっている。秩序状態のクレーム戦略学習において暗黙のルールとしての戦略が異なっているが、秩序状態ではクレームの数は少ないため行為だけ見てみるとどちらの戦略も同じである。アンケートによるとルールに従っていると答えた被験者も、実際は戦略的にクレームをつけていなかった可能性も考えられる。そのためこの二つの戦略は近いものと考えられる。また、エージェントの勝ちカード選択学習では戦略なしとなっているが、4割のエージェントは自分を親にする戦略に収束しつつあり、最も多い戦略となっている。これらにより、エージェントは人間と同様な暗黙のルールの獲得を行っているといえる。

また、各実験条件にて、Online BARNGA終了時に収束している戦略の人数を、勝者決定の戦略とクレーム戦略の分けてまとめたものを表6と表7に示す。

勝者決定戦略においては、条件Cによる影響が大きい。これは、学習期間が短くなるとルール獲得に弊害が生じ、ルール獲得が困難になり最終的に自分を勝者にする戦略が生じたと考えられる。一方、クレーム戦略においては、条件bによる影響が大きい。これはクレームの権力が小さくなったため、相手に罰則を与えるという意味でのクレームではなくなってきたと考え

表6 Online BARNGA終了時の勝者決定の戦略の人数

条件	ルール	自分	その他
Aa	11	1	3
Ab	13	2	0
Ba	10	1	4
Ca	5	1	9
Bb	13	1	1
Cb	6	7	2

表7 Online BARNGA終了時のクレームの戦略の人数

条件	ルール	クレーム	従う
Aa	10	3	2
Ab	5	0	10
Ba	12	3	0
Ca	11	4	0
Bb	2	0	13
Cb	5	9	1

られる。そのため、クレームは、ボーナスポイントを得るために付けることに変化していったと考えられる。

5.3 考察

人間の暗黙のルールを観察した知見を用いて、暗黙のルールを獲得するエージェントを作成した。その結果、人間と同様な社会的スキル：暗黙のルールの獲得を実現することができた。しかし、今回出現した行動は、エージェントの変数(学習率など)の設定に依存し

ているところが大きい。さらに、これらの設定は同一で、エージェントはホモジニアスである。人間同士の実験より、実際の人間は個体差が大きいことも分かっている。人間の個体差の観察を深く行えば、エージェントの個性の設計も可能になるであろう。そして、エージェントをヘテロジニアスとして設計をして実験を行えば、さらに人間に近い結果が得られると考えられる。

また、今回提案したエージェントは暗黙のルールの獲得ができたが、実際の人間は、今回実装した暗黙のルールの獲得以外の社会的な適応・学習能力を備えている。例えば、相手の裏をかくや腹を探るといった能力である。これらの能力は、個体の戦略として集団適応に大きく関わっていると考えられる。暗黙のルールの獲得と合わせて集団適応を設計すると、より興味深いエージェントを設計できるであろう。

6. おわりに

人間の暗黙のルール観察をおこない、暗黙のルール獲得のための能力を検討した。その知見を用いて、3つの状態に「気づき」と「振舞い」により遷移でき、規則の変化と戦略を別々に学習できるような、エージェントの設計を行った。このエージェントは暗黙のルールの獲得をすることができた。また、人間の実験とエージェントの実験を比較すると、近い戦略に収束した。これより、暗黙のルールを獲得するスキルを備えた、集団に適応をする社会的エージェントの設計ができたと言える。さらに、人間は「暗黙のルールの獲得」だけでなく「相手の腹を探る」や「相手の裏をかく」といった行動も行っている。これらを考慮するとより興味深い社会的なエージェントが実現できるであろう。

また、今回用いた環境であるBARNGAでは社会に存在する暗黙のルールを非常に単純化したものを用いている。そのため人間の生活におけるすべての暗黙のルールを再現できているとは言えない。今後実社会での活動を検討するためには、実際の社会での暗黙のルールに照らし合わせた環境を作り出し、今回提案したエージェントの実用可能性を考察をしていく必要がある。

今後、人間社会における集団適応エージェントが実現すると、人間に飽きないインタラクションを提供できるかもしれない。また、人間の集団の安定する状態を、提案エージェントを入れることにより変化させることができ、望みの状態に収束させることが可能になるかもしれない。

提案したエージェントはHAI研究において集団適応という新しい方向性を与えるものとなるであろう。

参考文献

- [1] 山田誠二. 人とロボットの〈間〉をデザインする. 東京電機大学出版局, 2007.
- [2] 片桐恭弘. ロボットの社会的知能. 情報処理, Vol.44, No.12, pp.1233-1238, 2003.
- [3] 中島宏, 森嶋泰則, 山田亮太, S.Brave, H. Maldonado, C. Nass, 川路茂保. 人間-機械協調システムにおける社会的知性. 人工知能学会論文誌, Vol.19, No.3, pp.184-196, 2004.
- [4] 片上大輔, 大村英史, 安村禎明, 新田克己. 社会的インタラクションに基づくマルチユーザ学習エージェント (MULA). 日本知能情報ファジィ学会誌, Vol.17, No.3, pp.340-350, 2005.
- [5] 大村英史, 片上大輔, 新田克己. クラスタリングを用いたマルチユーザラーニングエージェント (MULA-C). 人工知能学会論文誌, Vol.22, No.6, pp.621-630, 2007.
- [6] 大村英史, 片上大輔, 新田克己. 社会的エージェントのための人間の社会スキルの分析と検討. 第24回ファジィシステムシンポジウム, 2008.
- [7] 大村英史, 片上大輔, 新田克己, 野澤孝之, 近藤敏之. 人間の暗黙のルール獲得時の脳活動測定. 第17回北信越支部シンポジウム, 第32回関東支部ワークショップ, 第3回人間共生システム研究会, 2008.
- [8] 藤川吉美. 合意形成論. 成文堂, 2008.
- [9] 山本七平. 「空気」の研究. 文春文庫, 1983.
- [10] D. Goleman. *Emotional Intelligence : Why It Can Matter More Than IQ*. Bantam Dell Pub Group, 1995.
- [11] SST. <http://www.jasst.net/>.
- [12] 村井俊哉. 社会化した脳. エクスナレッジ, 2007.
- [13] 磯貝芳郎. 人間と集団・社会. 勁草書房, 1986.
- [14] 柴田克成, 上田雅英, 伊藤宏司. 強化学習による個性・社会性の発現・分化モデル. 計測自動制御学会論文誌, Vol.39, No.5, pp.494-502, 2003.
- [15] S. Thiagarajan. *BARNGA : A Simulation Game on Cultural Clashes*. Intercultural Press Inc., 2006.
- [16] 新井潔, 兼田敏之, 加藤文俊, 中村美枝子, 高木晴夫, 木嶋恭一, 出口弘. ゲームシミュレーション. 日科技連, 1998.
- [17] 山田祐二, 高玉圭樹. エージェントが集団へ適応するための評価軸の一考察: barnga における訂正行動の観察. 計測自動制御学会システム・情報部門学術講演会 2004, pp.19-24, 2004.
- [18] M. Jueptner and C. Weiller. Does measurement of regional cerebral blood flow reects synaptic activity?-implications for pet and fmri. *Neuroimage*, Vol.2, pp.148-156, 1995.
- [19] R. S. Sutton and A. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, The MIT Press, 1998.
- [20] S. W. Wilson. Classier tness based on accuracy. *Evolutionary Computation*, Vol.3, No.2, pp.149-178, 1995.

(2009年3月1日 受付)

(2009年8月11日 採録)

[問い合わせ先]

〒226-8503 神奈川県横浜市緑区長津田町 4259-J2-53

東京工業大学

大村 英史

TEL: 045-924-5218

FAX: 045-924-5218

E-mail: ohmura@ntt.dis.titech.ac.jp

著者紹介

おおむら ひでふみ
大村 英史 [非会員]

2002年東京農工大学工学部機械システム工学科卒業。2009年東京工業大学大学院博士課程修了。博士(工学)。同年東京工業大学大学院総合理工学研究科助手。2006年ハートフォードシャー大学客員研究員およびチューリヒ大学客員研究員。2007年東京工業大学大学院総合理工学研究科助教。現在に至る。人間・エージェントの集団適応に興味を持つ。人工知能学会の会員。

かたがみ だいすけ
片上 大輔 [正会員]

2002年東京工業大学大学院博士後期課程修了。博士(工学)。同年東京工業大学大学院総合理工学研究科助手。2006年ハートフォードシャー大学客員研究員およびチューリヒ大学客員研究員。2007年東京工業大学大学院総合理工学研究科助教。現在に至る。ヒューマンエージェントインタラクションに関する研究に従事し、知的エージェントを用いた集団適応ならびに関係性の可視化に興味を持つ。人工知能学会、日本ロボット学会、IEEE各会員。

にった かつみ
新田 克己 [非会員]

1975年東京工業大学工学部卒業。1980年同大学院博士課程修了。工学博士。1980～95年電子技術総合研究所に勤務。その間、1989～94年の間に(財)新世代コンピュータ技術開発機構に出向。1995年から東京工業大学大学院に勤務。現在に至る。主な研究内容は、論理プログラミング、エキスパートシステム、マルチエージェント、ヒューマンインタフェースなど。特に、法律分野への応用に興味をもつ。

のざわ たかゆき
野澤 孝之 [非会員]

2002年東京工業大学大学院総合理工学研究科博士後期課程修了。博士(理学)。同年京都産業大学特定研究員。2003年大学評価・学位授与機構評価研究部助手(2006年より助教)。2007年東京農工大学工学府特任助教。現在に至る。複雑系の分析手法とシミュレーション、脳活動はじめ生体計測データの分析・応用に関する研究に従事。電子情報通信学会、情報処理学会、日本数理生物学会の会員。

こんどう としゆき
近藤 敏之 [非会員]

1999年名古屋大学大学院工学研究科博士後期課程修了。博士(工学)。同年カリフォルニア大学サンディエゴ校客員研究員。2000年東京工業大学大学院総合理工学研究科知能システム科学専攻助手。2006年東京農工大学大学院共生科学技術研究院講師。2007年同准教授。現在に至る。2004年計測自動制御学会論文賞受賞。運動学習の実験的解析ならびに複雑適応系の工学応用に関する研究に従事。計測自動制御学会、日本ロボット学会、IEEEなどの会員。

Development of Social Adaptive Agents in Simulation Game of Cross- Cultural Experience by

Hidefumi OHMURA, Daiske KATAGAMI, Katsumi NITTA, Takayuki NOZAWA and
Toshiyuki KONDO

Abstract :

Recently, researches of Human-Agent Interaction (HAI) are popular toward autonomous agents to act and cope in various human societies. If agents adapt to humans' society, they should have social skills. These can't be dealt with through conventional methodology of HAI. In this paper, we introduce how developing a social adaptive agent. There are many rules in societies. Therefore, social agents should have a social skill which is obtaining social rules. We develop agents adapting communities as social adaptive agents. We improve some functions from a simulation game : *Online BARNGA* from *BARNGA* to observe humans' behavior. We observe humans' behavior in *Online BARNGA*. As a result, we find that humans have transitions of three inner states with "notice" and "behavior" to get implicit rules, and rules fall into two categories : a byelaw and an ethic. In *Online BARNGA*, a byelaw is a rule of card game, an ethic is a strategy. We analyzed the brain activities of humans obtaining implicit rules by the functional near infrared spectroscopy technique. As a result, humans' blood flow was decreased at complained by others, and increased at thinking, and it was related to learning. Based on the above results, we develop agents with transitions of three inner states with "notice" and "behavior" and two type learning modules. Furthermore, we analyze agents' behavior in *Online BARNGA*, and compare humans' that. As a result, we confirm that the agent adapts socially as if it were human.

Keywords : HAI, social adaptive agent, adaptation to groups

Contact Address : **Hidefumi OHMURA**

Tokyo Institute of Technology

4259 Nagatsuta-cho, Midori-ku, Yokohama, 226-8503, JAPAN

TEL : 045-924-5218

FAX : 045-924-5218

E-mail : ohmura@ntt.dis.titech.ac.jp