

「相手を読む」戦略の獲得—標準ゲームの提案

The acquisition of the strategy to “read others” – A proposal of a standard problem

伊藤 昭
Akira Ito後藤 強
Tsuyoshi Goto水野将史
Masafumi Mizuno寺田和憲
Kazunori Terada

岐阜大学工学部

Faculty of Engineering, Gifu University

To read other's mind/action is an essential ability in interacting with others. So far, what it is to read other's mind/action is not much investigated. We argue that what is missing in computer is an ability to read others and plan an strategy to *outwit* others in competitive environments. Hence, to develop an ability to mind-read others, a necessity for a standard problem is proposed, and its requirement is discussed. A kind of dungeon game is proposed as a candidate for a standard problem. Lastly, our long-range goal of mind-reading is explained.

1. はじめに

相手(の心, 行動)を読むということは, 我々が社会で生きていくにはなくてはならない技能である. 事実我々は, 相手の心を読んで適切に対応することで, 社会生活を営むことができる. スベルベル [1]によれば, 相手の心を読むということはコミュニケーション能力の基本をなすものであり, 言葉によるコミュニケーションも心を読む(相手の思考を推論する)能力の上に初めて成り立つという.

心を読むということをアルゴリズムとして分析することは, これまで殆どなされてこなかった. というのも, これまで心を読むことに関して多くの研究がなされてきたのだが, その説明の過程で常識的知識が必要に応じて援用されるため, その推論に必要な十分な知識が何かが明確にされなかった. 我々がここでアルゴリズムと呼ぶのは, 使われる知識と仮定とを明確にした上で, 手続きの形で推論を記述することである.

たとえば, スベルベルはコミュニケーションのための推論の方法を関連性理論として定式化し, その具体例を興味深い例を用いて説明している. しかしながら, その分析は本質を突いたものであっても, 残念ながら計算機プログラムとして実装できるようなアルゴリズムにはなっていない. というのも, それらは人の常識という背景知識の上で説明が行われており, また上記のアルゴリズムという観点から言えば十分に定義されていない概念がしばしば使われているからである. たとえば, 関連性の計算式とは何か, 我々はそれを共有しているのか, 共有しているのだとすればどのようにして共有するに至ったのか, など解明すべきことは多い.

我々もやはり, コミュニケーションにおける「心を読むこと」の解明を最終目標と考えているが, ここでは研究の方向を絞って, もう少し機械的な「相手(の心, 行動)を読む」ことのアルゴリズムを取り上げてみたいと思う. もし相手を読むことの複雑さがそこに用いられている知識の複雑さから来るのであれば, そこに用いられている推論に何ら特殊なことがないとするれば, 我々のアプローチは新しい知見にはいたらない. 我々は, 相手を読むことが, 一般の推論とは違った特殊性を持っていると考えており, その特徴を解明することが, 心を読むことの解明につながると信じている [2].

2. 対戦ゲームと「相手を読むこと」

「(相手を)読む」という言葉でまず思い浮かぶのは, 囲碁や将棋などの対戦ゲームで, 相手の指し手を予想しその対策を考えることであろう. 「相手の手を読む」, 「読みが深い」などの言葉は本来そのような意味からの日常生活への転用である.

しかしながら現実には, 対戦ゲームにおいて「手を読む」ということは多くの場合最善手の探索を意味し, 「相手の指す手を予測する」という意味で相手の手を読むことはあまり推奨されていない. むしろ場合によっては, そのような手は「嵌手(はめて)」と呼ばれて蔑まれることさえある. というのも, 囲碁や将棋のような完全情報零和ゲーム(手の決定に必要な情報はすべて知られており, かつ相手の利益と自己の利益とが完全に対立するゲーム)では, 必ずしも相手の行動を予測する必要はなく, 相手が最強の手を指すものと仮定して, その場合における最善手を探索する(いわゆるミニマックス戦略)ことができれば十分だからである. この場合, 最善手は, 相手が誰であるかにかかわらず盤面から決まる手となる. もし最善手が見つかるのであれば, 相手のモデルは必要がなく, また最善手を指し続けることは極めて合理的な戦略のように見える.

しかしながら, 現実には必ずしもそうではない. たとえば, そのような最善手を指しては勝ち目がないと思われたとき, 人は「まぎれ」を求めて最善手から外れた手を指すことがある(勝負手, ときには嵌手と呼ばれる). 当然, それは相手に最善手を出された場合不利になるため, その採用のためには適切な相手モデルが不可欠である. このような最善手から外れた手を指すのは特殊な場合であり, 普通はそのような手段を採るべきではないと言われるかも知れない. しかしながら, 不利な状況を打開するためにはプロ棋士も「勝負手」を指す.

この問題に関して, IBMの開発したチェスマシン Deep Blue とチェス世界チャンピオンのガスパロフとの対戦には示唆深いものがある. この事件は, ついに計算機(人工知能)が人の知能に勝利した(もちろん, チェスという限られた領域ではあるが)ということでき大きく持ち囃されたものである. 面白いことに, 飯田 [3]によれば, この戦いではガスパロフは心理作戦で負けたとされている. すなわち, Deep Blue にはある種の癖(評価関数の偏り)が意図的に作ってあり, ガスパロフがその弱点を突こうとしたとき, それに対する対策が Deep Blue 側に用意されていたというのである.

これは世間の言うように, 本当に計算機の勝利と言えるのであろうか. 確かに試合は Deep Blue 側の勝利であった. しかしそれは, Deep Blue 開発チームの勝利であったとしても, 「計

算機」自身の勝利とは見做し難い。実際、Deep Blue の最終調整は当然ながら人手で行われている。このとき、ガスパロフについてはあらゆる情報が収集/分析されている一方、Deep Blue についてはガスパロフはその実体を殆んど知らされていなかったという。とすれば、明らかにこの対戦は「相手を読む」という点からは公平な戦いではなかった。飯田も言うように、繰り返し対戦すればガスパロフが勝つ可能性は十分高くなるであろう。

計算機に欠けているものは、何であろうか。それは相手のモデルを作り、それに基づいて相手に読み勝てる戦略を(プログラム)自らが立てることである。残念ながらチェス程度の問題でも、このような「戦略」を立てる能力では計算機は人に遥かに及ばないのが現状である。自然言語の理解と異なり対戦ゲームは非常に論理的であり、このような能力の欠如をいわゆる「常識」の欠如ということでは言い訳することは許されない。今の計算機には、対戦ゲームにおいても不可欠な、相手を読むという能力が本質的に欠けているのである。

チェスゲームについて言えば、計算機のプログラムがいつでも自由に対戦できる形で公開され、人手による調整を経ないで勝ち続けられるようになったとき、はじめて人に勝つプログラムができたと言えるのではないかと私は思っている。

3. ゲーム理論

相手の行動を予測して自らの戦略を決めるという問題は、本来ゲーム理論が扱おうとした問題である。しかしながら不思議なことに、ゲーム理論においても相手のモデルに基づき相手の行動を予測するということは、その研究の中心的関心とはならなかった。ここでもやはり、「相手は合理的な行動主体である」ことが早い段階で仮定され、その仮定の下で最適な行動を求めることが研究の主目的とされた。もちろん、このような仮定のおかげで、様々な有用な定理、現象が発見されたわけであるが、相手を読むということの分析が放置されてしまったことも残念ながら事実である。

これに対して、社会学、動物行動学などでは、現実に人や動物が一見ゲーム理論に反した不合理な行動をとるということで、それを説明するためのさまざまな研究がなされてきた。たとえば、Axelrodらは繰り返し囚人のジレンマゲームを用いて、協調がどのようなときに生じるのか、どのような戦略が有利かについて、包括的な研究を行っている [5]。しかしながら、ここでも相手を読むことは結果として重視されなかった。というのも、彼らの一連の研究では、繰り返し囚人のジレンマゲームで良い得点を上げるのは、非常に単純なしっぺ返し(TFT:Tit For Tat)と呼ばれる戦略であったからである。この結果は不幸にして、ゲーム理論においては複雑な戦略を用いてもあまり良いことはない、と言う印象を多くの人に与えてしまった。

しかしこれは、本当にそうであろうか。実は、TFTが有利であったのは、たまたま周りのプログラムが良い(上品な)戦略ばかりであったからに過ぎないのではないか。もっと状況が複雑になれば、TFTのような単純なプログラムではなく、相手の手を読むことで様々な相手に対して適切な戦いをするプログラムが必要とさるであろう。またそのようなプログラムは、周りが上品なプログラムでもそこそこ良い性能を示し、複雑な環境下では平均として高得点をあげるのではないだろうか。実際、ALIFE V 国際会議で行われたノイズあり繰り返し囚人のジレンマゲームコンテストで、我々は相手を読むプログラムをエントリーして良い成績を納めている。

しかしながら、「相手を読む」プログラムの実験を包括的に行うためには、残念ながら繰り返し囚人のジレンマゲームは単純すぎるゲームである。そこでは、相手を読もうとしても利用できる情報は限られている。我々は、以前に「対戦履歴を公開するという条件下での次々と相手を変える囚人のジレンマゲーム」を提案し、そのような状況下では「相手を読む」戦略が有効なことを示した [4]。そこでは、対戦相手が様々な相手と対戦したときの対戦履歴が利用できるため、利用可能な情報は複雑であり、それに応じてより複雑な戦略が効果を発揮する余地が生じる。

4. 「相手を読む」標準課題の条件

ここでは、相手を読むことに焦点を当てた研究を行うために利用可能な標準課題の条件を検討する。

課題の条件の第一は、現在の技術水準でも、相手を読む能力の違いにより成績に差が出ること、また単純な戦略でないものが優位性を発揮できることである。ただし、ここでの能力とは、問題固有の技術ではなく相手の行動をどう予想するかという能力である。逆に問題固有の技術が計算機にとって難しいと、結果としてその技術の優劣が勝敗を左右することになり、相手を読むこと以前の段階で勝敗が決してしまうことになる。

たとえば、チェスや将棋、囲碁などの対戦ゲーム、RoboCup(シミュレーションリーグ)などでは、問題固有の技術が優劣を大きく左右する。すなわち現状では、(中途半端に)相手を読む(相手の行動を予測する)ことよりも、問題固有の技術(盤面評価、バスの技術)の改良に重点をおいた方が、勝率ははるかに上がるのである。もちろん、最終的にはそのような課題にこそ相手を読む能力が必要とされるのであろうが、現状では相手を読むことの有用性が技術的問題に隠されてしまう。

一方、じゃんけんなどのような単純なゲームでは、もし一方的に相手を読む(予想する)ことができれば圧倒的な優位に立つことができるが、読まれる側は確率的に行動を選択する(混合戦略)ことで、相手の読みを拒否することができる。したがって、このような単純なゲームでも、やはり相手を読む能力が実力を発揮する余地があまりないことになる。したがって、相手を読むことの標準課題としては、問題固有の技術が大きくものを言うほど複雑ではなく、またじゃんけんのように確率的戦略が容易に見つかるほど単純でもないものということになる。

課題のもう一つの条件は、対戦相手も対等な条件で競技に参加するということである。固定的な行動戦略を取っている相手の行動を読むことは、物理環境の学習と同じ課題であり、(もちろん重要な課題であるが)ここでの研究対象ではない。相手を読むことで一番難しい課題は、「相手も自分を読んでいる」ということに如何に対処するかということである。

もちろん、相手に計算能力で圧倒することができる行動主体であれば、相手がこちらを読んで戦略を変更する前に、相手を「変化しない」と見做してこちらの戦略を調整することが可能である。しかし良く考えると、一定以上の計算能力を持った行動主体同士では、このようなことは起りそうにもない。相手の行動を予測するオートマトンは、予測しようとする対象オートマトンより複雑度が指数関数的に優位(相手の複雑度を N とすると、自己の複雑度は 2^N 以上)でないといけなことが分かるからである。

5. マルチエージェント強化学習

相手がこちらの行動に応じて変化する以上、相手のモデルは事前に入手できる情報だけでは不十分であり、対戦中に入手できる情報を積極的に使わなくては、相手に打ち勝てる予測は不可能である。逆に、自己の戦略が固定的であれば、それが如何に優秀なものであっても、いずれは相手からのモデル化を許し、自己の立場を不利にすることになるであろう。このように、相手をモデル化することを「行動の中での学習」と捉えると、この課題は本質的にマルチエージェント強化学習として定式化できることがわかる。強化学習という言葉は、解くべき問題としても、またそれを解く手法としても使われているが、ここではこれを「問題としての強化学習」として定式化に利用する。

事前に解くべき問題のクラス（要素を指定する変数が含まれる）が事前に与えられる。その制約の中で、強化学習が定義される。エージェントは、時刻 t で状況 $S \in \Sigma$ を認知し、可能な選択肢 A から自己の行動 $a \in A$ を決定する。その結果報酬 r を受け取り、時刻 $t+1$ に移行する。課題は、時刻 t までに得られる情報を自由に使って、行動 a を選択し、事前に与えられた終了時刻 T までの総報酬量を最大化することである。なお S , r は、それ以前に選ばれた a 全てに依存しても良い（マルコフ性は仮定されない）。

事前に与えられる問題のクラスとは、状態 S , 報酬 r を定める規則である。実験は、問題のクラスから少数のインスタンスを生成して行われる。たとえば RoboCup シミュレーションは、参加チーム（クライアント）プログラムがどのような問題を解かなければならないかという視点から見れば、このパラダイムとして捉えられる。

6. 研究課題

標準課題を解くにあたって、

1. 良い成績を獲得するプログラムの人手による設計
 2. 良い成績を獲得するプログラムの計算機による自動獲得
 3. 人の用いる戦略とプログラムの戦略との比較
- のいずれを目指すのかで研究内容は異なる。

1. は、人手でもかまわないので強いプログラムを作ろうということである。ただし、「強い」という意味の定義は難しい。一つの考え方は、Axelrod のように、コンテストで出場チーム同士を競わせるものである。おそらく、相手に合わせて行動を調整する能力が必要とされるであろう。2. は、汎用的な学習器だけを留意して、行動の中で戦略を獲得させようというものである。実は我々は、1., 2. は長期的には一致すべきものと考えている。すなわち、自らを自動的に改変することのできるプログラムでないと、「相手を読む」ことの必要な課題では良い成績を得られないはずである。

3. は、1, 2 の越えるべき目標を指し示すとともに、本研究の最終目的でもある。もし人の用いる戦略が 1, 2 などよりも良い性能を示すとすれば、ここに新たな研究課題が生じる。人はどのようなアルゴリズムを用いて、そのような戦略を編み出すことができるのであろうか。そのような「戦略生成器」を獲得するために、どのような知識/情報が必要となるのであろうか。

前に述べたように、将棋や RoboCup のように複雑な課題では、人の戦略に勝てる計算機プログラムはいまだ存在しない。したがって、それらの課題は標準問題としての条件を満たしているように見える。しかしながら、人の戦略を分析しようとする、このような複雑な問題では人の常識の壁に阻まれて、「相手を読む」こと自体のアルゴリズムがなかなか明らかになって

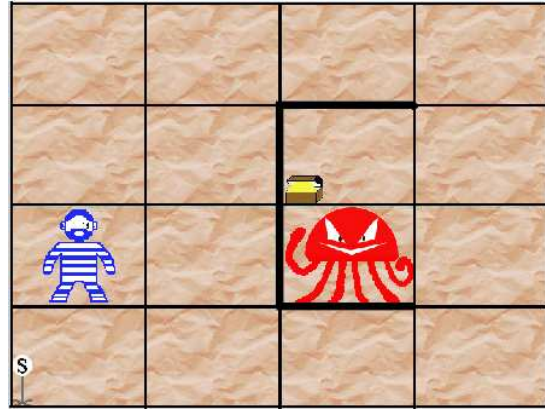


図 1: ダンジョンゲーム

こない。標準問題として比較的簡単な問題を用いることで、人における相手を読む戦略のアルゴリズムの分析が容易になる。また機械学習アルゴリズムは、十分な計算（探索）時間が与えられれば最適な戦略を生成できるはずである。ここでも比較的簡単な課題を用いることで、人の戦略と機械学習の獲得した戦略とを比較することが可能となる。

もしこの時、人の戦略の方が性能が良いとしたら、そこで我々ははじめて人の「相手を読む」戦略の本質にたどり着くことになる。機械学習の結果得られる戦略は、ある条件下では最適性を保証されているものである。人の戦略がそれに勝つことがあるとすれば、そこには何らかの根拠があり、その能力の獲得が計算機による「相手を読む」能力につながるはずだからである。もちろん、No Free Lunch Theory [6] の教えるところによれば、一般的な問題群に対して常に効率の良い探索方法はない、とすれば、人は何らかの意味で「偏った」モデルを用いて対象をモデル化しているはずである。我々はこのような「偏り」を「相手を読む」ことの本質として考えている。

7. ダンジョンゲーム

以下では話を具体的にするため、我々の考えている標準課題の例として、ダンジョンゲームを説明する。これは図 1 のような簡単なマップを持つダンジョンで、モンスターと海賊とが自己の得点を求めて行動する問題である。海賊、モンスターは、壁で遮られない限り上下左右に 1 マス動く、または現在の位置に留まることができる。海賊は、ゴール（宝の場所）へ着くとゴールポイント G を得てスタートへ戻される。また、モンスターと衝突すると衝突ポイント H をモンスターに奪われる。一方モンスターは、海賊と衝突すると衝突ポイント H を海賊から獲得する。モンスターは、ゴール、スタートには侵入できない。ゲームの目標は、海賊、モンスターともに、与えられた時間内に最大の得点を獲得することである。

一見単純なゲームに見えるが、それでも最適な行動を決めるのはなかなか難しい。モンスターの立場からは、海賊を追いかけても相手が逃げようとするれば、決して自力で捕まえることはできない。逆に海賊の立場からは、モンスターが宝の前に居座ると、衝突せずにゴールに着く方法はない、いずれにしても、この問題は相手の協力がなくては得点が得られない設定になっている。したがって、状況は囚人のジレンマゲームと似ている。

囚人のジレンマゲームとの違いは、海賊とモンスターとの非対称性である。この問題では囚人のジレンマゲームでのしっぺ返し戦略のように、常識的にみて「妥当な」戦略というのは

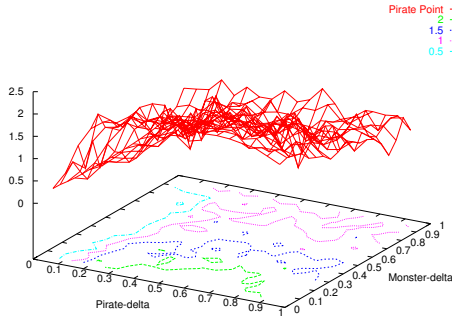


図 2: 海賊の平均得点

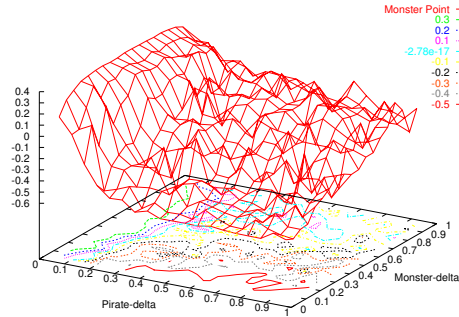


図 3: モンスターの平均得点

存在しない。しかしながら我々は、この問題が囚人のジレンマゲームと類似と気が付けば、短期的利益を追求する以外の「友好的」戦略を思いつくことができる。実際、海賊、モンスター全体として得点を得る方法は、海賊がゴールに行くことしかないのだから、モンスターはそれを完全に妨害してはいけない。また海賊もモンスターの協力をえるためには、相手にも適当な分け前を渡さなければならない。

一方、単純な強化学習エージェントからみれば、このような戦略の獲得はほとんど不可能なように見える。というのも、モンスターには海賊がゴールを目指しているということは観測からは知り得ないことであり、モンスターは「ゴール付近で待ち伏せすればよい」ということを推論するための知識を持ち合わせていないからである。

実は、このような戦略は機械的な強化学習によってもある程度は獲得可能である。図2, 図3は、海賊、モンスターの学習速度を様々に変化させたときの海賊、モンスターの平均得点である。細かい説明は省略するが、自己の学習速度が相手のそれを大きく上回るとき、自己のにとって有利な形で様々な戦略を獲得する。その結果、そのようなパラメタ領域では、自己の平均得点が大きくなる。また中間領域では、それに応じた様々な戦略が生成される。機械学習でも、長時間の経験によりまたまうまくいった状況を収集することにより、それなりの戦略を獲得することはできる。

では問題は解決されたのかというと、実は全くそうではない。人は多くの場合、極めて少数の事例から相手のモデルを生成することができる。たとえば、強化学習の標準的な枠組みでは、学習のためには全ての状態で可能な行動全てを多数回経験することが収束の条件となっている。事実上記のダンジョンゲームでは、極めて多数回の試行の後に自己に有利な戦略を獲得している。しかしながら、人は多くの問題で、少数回の経験で有効な戦略をつくることができる。この差は、残念ながら歴然としている。

8. 機械は統計的予測を越えられるか

機械が外部環境を学習する手法は、全て統計的学習にその基礎をおいている。すなわち、それらは全て、様々な事象の統計的性質(相関)を求めることで、未来を予測しようとするものである。HMMとか、ニューラルネット、Q学習などの機械学習は全てこの枠組みに含まれる。

我々が外部環境を学習するときには、統計的学習を用いることも多いが、それ以外の予測をしばしば行っている(ように見える)。たとえば、Dennettの設計スタンス、意図スタンス

[7]では、製作者、または対象自身の「意図」を考えることで、その行動を予測する。具体的には、まず説明対象を「意図」を持っている物と見なす。そうして、可能な意図を仮定し、それにしたがって行動を予測し、実際の行動と比較し、検証する。検証されれば、以後はその意図を用いて相手をモデル化することで、遥かに容易に相手の行動予測が可能となる。

実際にダンジョンゲームでは、統計的予測と意図スタンスとの違いはどのように現れるのだろうか。たとえば、海賊が特定の状況ではある行動パターンを取る可能性が高いという統計的情報を収集することは、時間をかければ可能となる。しかしながら、海賊がある地点(ゴール)に行こうとしているということを、観測だけから判断するのは難しい。しかし、一旦そのことを理解してしまえば、海賊の行動を予測することは統計的予測だけに頼るよりも遥かに高速で、かつ精度が高いものとなる。人はこのようなことを容易にやっけるのである。

したがって、「相手を読む」研究の最終目標は、そのような能力が計算機によりどのようにして、またどのような条件で獲得できるのかを明らかにし、計算機にも統計的学習を越える学習能力を持たせることである。ここでは、そのような試みの実験を行うための「標準課題」についての条件と、その候補問題の提案を行った。

参考文献

- [1] Sperber, D. and Wilson, D.: *Relevance: Communication and Cognition*, Oxford, Basil Blackwell, 1986.
- [2] 伊藤 昭:「心を読む能力の創発 —マルチプレイヤー囚人のジレンマゲーム」, 認知科学, Vol.6 No.2, 1999.
- [3] 飯田 弘之:「コンピュータは名人を超えられるか」, 岩波書店, 2003.
- [4] 伊藤 昭, 矢野 博之:「利己的なエージェントの社会におけるつきあい方戦略の進化」, 情処論, Vol.38, No.5, pp.944-952, 1997.
- [5] Axelrod, R.: *The Evolution of Cooperation*, Basic Books Inc., (1984), 松田裕之訳, つきあい方の科学, HBJ 出版局, 1987.
- [6] Wolpert, D.H.: "The relationship between PAC, the statistical physics framework, the Bayesian framework, and the VC framework", Wolpert, D.H.(ed.), *The mathematics of Generalisation*, Addison-Wesley, Reading, MA, 1995.
- [7] Denett, D. C., *Kinds of Minds*, Basic Books, 1996.