E-020

対戦型不完全情報ゲームにおける人の戦略モデル
A model of the human strategies in an multi-player imperfect-information game

水野 将史 † 伊藤 昭 ‡
Masafumi Mizuno Akira ito

1. はじめに

お互いが学習するエージェントによる対戦ゲームでは、対戦相手が用いている戦略を「読み」、相手に勝る戦略をとることができれば相手を負すことが出来る。我々はこれまで、対戦型マルチエージェント問題に取り組み、学習する対戦プログラムの研究を行ってきた。その過程で、主として物理的（静的）環境を学習するために開発されてきた強化学習 (統計学習) の不十分さを認識し、強い学習プログラムを開発するためには、人のとる戦略（思考方法）の再検討が必要となると考えるようになった。

本発表では、我々が解こうとしている問題を提出し、このような問題に対してこれまでの学習アルゴリズムの不十分さを指摘する。次に、実際人同士や人対コンピュータの対戦を行い、人がどのような戦略を用いて問題を解くのかを実験的に調べ、人の取る戦略のモデル化を行う。また、このような問題を解くに際しての、人の学習と機械の学習との比較を行い、強い学習プログラムを開発するために必要な機能について検討する指針とする。

お互いが学習するエージェントによる対戦ゲームは、不完全情報ゲームとして定式化できる。不完全情報ゲームの研究は、単にこのようなパズル的なゲームを解くことが目的ではない。日常の対人（対エージェント）インタラクションにおいて常に遭遇する利害が対立する中での行動（意思）決定は、本質的に不完全情報ゲームと等価であり、計算機が不完全情報ゲームを解けることは、人間らしい思考をするコンピュータを開発する為に必須の技術である。

2. ダンジョンゲーム

以下では、不完全情報ゲームとしてダンジョンゲームを取り上げ、議論を展開する。

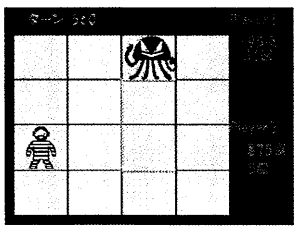

図 1: ダンジョンゲーム

ダンジョンゲームとは、図 1 のような 4 × 4 マスの空間で 2 種類のエージェントが対戦するゲームである。2 つのエージェントのうち Player(図 1 で人のようなもの) は相手に捕まらずに Goal に向かうことを目的とし、Monster(図 1 でタコのようなもの) は相手を捕まえることを目的とする。

†岐阜大学大学院工学研究科
‡岐阜大学工学部

Start は図 1 の左下、Goal は図 1 の中央右上、また、Goal の周りは壁のようになっていて、どちらのエージェントも通り抜けることは出来ない。Player のスタート位置は図 1 の左下であり、Monster のスタート位置はランダムに決定される。また、Player は Goal に到達すると、次のステップでどのような行動を選択しても Start に戻される。Monster は Goal に侵入することはできない。

Player と Monster は各ステップにおいて上下左右に移動するか、またはその場に留まるかの 5 種類の行動から一つを選択する。Player と Monster が行動を選択し終えたらそれぞれに次の位置を教え、報酬を与える。Player と Monster が得られる情報は現在の自己と相手の位置と自己の報酬だけである。

行動の結果各エージェントに与えられる得点を表 1 に示す。

	ゴール到達	捕まった時	一歩動く
Player	30	-15	-1
Monster	-	15	-1

表 1: エージェントの得点

なお、両者とも相手の点数は関係なく、自己の得点を上げることを目標とする。

3. ダンジョンゲームの戦略

我々はこれまで、このダンジョンゲームを学習するプログラムがどのようなに解くのか、すなわち学習プログラム同士を対戦させることで、どのような戦略が獲得されるのかを計算機シミュレーションにより調べてきた。その過程で我々が得た学習プログラムに対する「不満」が、本研究のモチベーションである。すなわち、学習プログラムは我々が当然強いプログラムが備えるべき戦略を獲得できないのである。

実際、このゲームを学習プログラム同士で対戦させると、次のような戦略の学習が行われる。図 2 に、学習過程での Player・Monster の得点を示す。(横軸はターン数である。) 最初は Player, Monster ともランダムに行動するが、捕獲 (衝突) により得点の移動があることを教えられると、Player は Monster を避けるように、また Monster は Player を追いかけるように学習する。しかしながら、同じ速度で行動することの出来る Player と Monster では、Player が回避行動をとる限り Monster は Player を捕獲することは出来ない (ターン t=2000)。

その後少し遅れて、Player がゴールに行くことを学習する (t=10,000)。しかしながら、Player がゴールに行く行動を繰り返すと、Monster は Player の経路上で待ち伏せすることを学習する (t=100,000)。このため Monster の得点は増加し、Player の得点は減少する。しかしながら、Monster が Player を捕獲することを繰り返すと、

Playerはもはや危険を犯してまでGoalを目指さなくなり、ついには全くMonsterはPlayerを捕獲することができなくなる(t=350,000)。

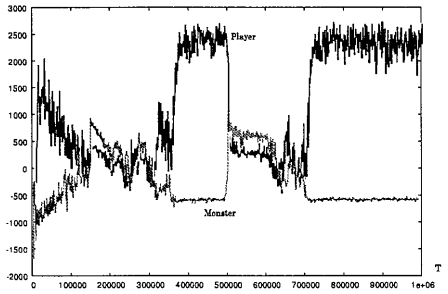

図2: Q学習エージェント同士の対戦結果

この時点で待ち伏せ利益のなくなったMonsterはPlayerを探して彷徨し始め、PlayerはMonsterのいなくなった経路を通ってゴールに到達する。最初はMonsterを避けながらGoalを目指していたPlayerは、Monsterが経路上に現れないことから、そのうちMonsterを気にせずに行動するようになる。そうして、あるとき突然Monsterが経路上に出現し(t=500,000)また、先ほどと同じプロセスが繰り返される。

3.1 最強の戦略とは

学習プログラムの動きを見て、人はこのプログラムがそれほど利口ではないと思ってしまう。Playerは絶対にMonsterに捕まらないようにすることで、有利な状態(t=400,000前後のような)を維持することができる。一方、Monsterは一旦Playerが来なくなっても、我慢強くGoal付近で待ち伏せを続ければ、いずれはPlayerの方から来てくれるはずである。

では、それらの戦略が両者にとっての最善戦略かというと、そうではない。両者が上の戦略をとり続ければ、お互いににらみ合うだけで両者とも得点を得られないのである。このゲームにおいて単純に最適であると言える戦略は見つからないのである。

もう少し数学的にゲームの構造を分析してみると、両者の総得点をプラスにするためには、PlayerがGoalに行くしかない、ということがわかる。したがって、戦略はPlayerは捕まってもよいからGoalへ行くという戦略をとり、MonsterはPlayerを時々捕まえるという戦略をとることが必要である。すなわち両者が協力して両方がある程度の利益を得ることで妥協する協調戦略が必要なのである。

実は、これはゲーム理論では交渉問題として知られている問題である。交渉問題とは2人のエージェントがそれぞれ独立して最良の行動を取るよりも、2人が協力して協調行動を取ったほうが有利であるが、その妥協点をめぐって争わねばならない問題である。我々はこの問題を過去の履歴を用いる学習器に解かせることで、自己利益しか考えてないエージェント同士であっても、必要に迫られれば他エージェントとの協調行動が可能なことを確認した。

しかしながら、ダンジョンゲームに戻って考えると、これが交渉問題と等価であるということは自明ではない。学習器はどのようにして、問題の本質を抽出し、解決すれば良いのだろうか、と考えたとき、人でもこの問題を(ゲーム理論の知識なしに)解けるのだろうかという疑問が生じる。そこで、我々は様々な条件の下でこの問題を人に解かせることで、人はこの問題をどうモデル化し、どのように解いているのかを調査することとした。この知見は、今後より良い(強い)プログラムを開発する上において非常に重要になるはずのものである。

4. 被験者による対戦実験

以下に、我々が計画している実験計画を述べる。

始めに、人が学習プログラムを相手にダンジョンゲームを対戦した場合どのような結果が得られるか実験を行う。対戦相手には基本的なQ学習のアルゴリズムを適応する。まずは人がMonster側を操作し、学習するプログラムのPlayer側と対戦する。この実験では被験者に何を教示するか重要であるが、今回は対戦ゲームである事、自分の動かすキャラクターとその操作方法、自身の得点表示、およびゲームの目的がより多くの得点を得ることあると教示する。また、キャラクターはコンピュータが動かしていることも併せて教示する。これにより、人が相手キャラクタの行動をどのようにモデル化し、どのような作戦を立てるのかを調査、分析する。

つぎに、人と人がそれぞれPlayer、Monsterを動かして対戦する実験を行う。この時、対戦相手はそれぞれ見えない場所で対戦してもらう。実験前の教示は「人vsコンピュータ」の時と同様とするが、相手が学習する(賢い)コンピュータであると嘘の教示をするグループと相手が人であると教示するグループに分けて実験を行う。教示により作戦に違いが出るかどうかで、人の相手モデル化の方法を探る手がかりとする。また、事前に「人vsコンピュータ」の実験を行っている人に対して、「人vs人」の実験も行ってもらう、既に得たゲーム知識をどのように活かして作戦を立てるかを調べ考察する。

5. 人はどのような行動をとるのか

このようなゲームで人が戦略を考えるとき、まずゲームモデルの構築を行う、すなわち、どうすると点を得られるのかを検討すると思われる。多くの人は、いままでの経験から、このようなゲームの全体を把握するのにそれほど時間を要さないであろう。それよりも重要なのが、相手のモデル化である。人はどのように相手をモデル化しているのだろうか？それは、まず相手がどのようなものであるかという教示によって変化すると思われる。もともと、相手が人であると教示されていれば、人は相手も最大限に考えられた行動をすると思う。コンピュータと教示されていれば、まずは相手がどのように動くかを見極めようとする。

人は長期的な利益と短期的な利益の両者を互いにうまく利用する能力に長けている。この部分の戦略をコンピュータにうまく組み込む事が出来れば、人間らしいより良い(強い)プログラムを考え出すことが出来ると思われる。