

協力ゲーム Hanabi における相手の思考時間を戦略の指標に加えたエージェントの開発と評価

佐藤 栄介 加藤 拓也 大澤 博隆
筑波大学

概要：人間と協力して課題を達成するための人工知能は、人間の思考に合わせた行動をとる必要がある。協力ゲームの一つとしてカードゲーム「Hanabi」において、自己推定、心理学、コミュニケーション理論等さまざまな方面から研究が進められている。本研究では思考時間に関する仮説に基づき人間プレイヤーの思考時間の長短から人間側の AI の戦略に対する理解度を推測し、その結果に応じて自分の手札を推定する時の推定信頼度を変更する手法を用いたエージェントを開発した。また、開発したエージェントと従来のエージェントとの得点及び人間プレイヤーに与える印象の違いを調査した。結果、思考時間に応じて推定の信頼度を変えることにより得点や推定の成功率に有意差は見られなかったが、思考時間に関する仮説は妥当であることがわかった。また、思考時間に応じて推定の信頼度を変えるエージェントは従来のエージェントに比べてゲームに慣れておらず意思を持っていないという印象を人に与えることがわかった。

キーワード：Hanabi AI エージェント 協力ゲーム 思考時間

Development and evaluation of the game agent to add opponent thinking time as indicator of strategy in Cooperative game Hanabi

EISUKE SATO^{†1} TAKUYA KATO^{†1} HIROTAKA OSAWA^{†1}
University of Tsukuba

Abstract: AI to achieve the task cooperating with human have to act according to human thought. The research is advanced using cooperating game 'Hanabi' from various aspects such as self-estimation, psychology, and communication theory. In this research, we developed an agent that has the strategy to guess the degree of understanding of AI's strategy on human player side from the length of thinking time of the human player based on the hypothesis thought time and change the estimation reliability when estimating AI's own hand according to the result of it. We also examined the difference about score and impression given to human between this agent and the conventional agent. As a result, we found there was no significant difference in the score and the success rate of the estimation by changing the reliability of the estimation according to the thinking time but the hypothesis about the thought time was valid. We also found that the agent that change estimation reliability depending on opponent's thinking time impresses human as less used to the game and less intentional than the conventional agent

Keywords: Hanabi, AI, Cooperative game, Thinking time

1. 序論

現在、協力ゲームをおこなうための AI が注目されている。人工知能でゲームをプレイする分野において、人間に勝つことを目標としたエージェントは将棋や囲碁などで開発されており、その目標は達成されつつある。その一方で、人間と協力して一つの課題を達成するエージェントはまだ発展途上である。このようなエージェントが人間に勝つことを目標としたエージェントと異なる部分は人間の思考に合わせて人間に近い行動をとらなければならないところである。よって、このようなエージェントの開発は人工知能の人間らしさの獲得および人間の知能の人間らしさの解明において有意義なものであると考えられる。

協力ゲームをプレイする AI に関する研究の一つとして協力カードゲーム「Hanabi」を用いた研究が挙げられる。このゲームの大きな特徴は、相手の手札のみを見ることができて自分の手札はわからないことと、相手への情報提供がトークンによって制限されていることである。我々は Hanabi において他者の視点と行動をシミュレートして、自分の状態を推測するエージェントを開発した[1]。また、Eger らは心理学の手法とコミュニケーション理論から、人間がエージェントに対して何か意図を持って行動していると感じた時、人間はそのエージェントの能力を高く評価し好感を抱くということを、エージェントと人間で Hanabi をプレイさせることにより示した[2]。

ただ、既存の研究では人間側の思考時間という情報は

^{†1} 筑波大学
University of Tsukuba

Hanabi をプレイするエージェントの戦略においてまだまだ利用されていない。思考時間は人間プレイヤーのゲーム及び相手に対する理解度の指標として重要なものであるため、これを利用しなければ人間と協力できるエージェントの開発は困難であると考えられる。

本研究では人工知能を用いた **Hanabi** における人間との協りに適した戦略を検討する。エージェントが協力ゲームをプレイする場合の戦略は協力相手の思考に合わせたものであるため、人間とプレイする場合には相手が自分のとった行動の意図を理解してくれることが前提になっている。相手がエージェントの行動の意図を理解していないと、お互いの行動がうまくかみ合わず、課題の失敗につながる。相手がエージェントの戦略を理解しているならば推定の信頼度を上げて推定を積極的に利用して行動するべきであり、逆に相手がエージェントの戦略をあまり理解していないならばエージェントは推定の信頼度を下げて確定情報を重視して行動するべきである。そのため、相手の理解度の推定はエージェントが戦略を決めるにあたって重要な問題である。

本研究ではエージェントの戦略に対するプレイヤーの理解度をプレイヤーが行動を選択するまでにかかる時間から判断することを試みた。また、実際に人間にこのエージェントと **Hanabi** をプレイさせることによって推定の信頼度を変更しない場合に比べて得点および人間のエージェントに対する評価がどのように変化するか調査した。

本論分の構成は以下のとおりである。2 章では **Hanabi** のルールを説明し、3 章では従来の自己推定戦略のアルゴリズムと本研究で実装したアルゴリズムとの違いを述べる。4 章では評価実験の方法を説明し、5 章ではその結果と考察を述べ、6 章に結論を示す。

2. Hanabi のルール

本研究で用いるゲーム「**Hanabi**」のルールを以下に示す。

Hanabi は協力型のカードゲームであり、通常 2～5 人で行う。今回の研究では 2 人でプレイする時のみを扱う。**Hanabi** は 50 枚のカードと 8 枚の情報トークンを用いて行うゲームである 50 枚のカードにはそれぞれ色と数字が設定されており、色は白、赤、青、黄、緑の 5 色、数字は 1～5 の 5 種類である。カードは各色 10 枚ずつ存在しその数字の内訳は 1 が 3 枚、2～4 が 2 枚、5 が 1 枚である。このゲームの最終目的は盤面に各色のカードを 1～5 まで数字の昇順に並べていきすべての色で 5 までカードを並べることである。ゲーム開始時、プレイヤーにはそれぞれ 5 枚のカードが手札として配られる。この時、自分の手札の情報は見ることができず、相手の手札の情報のみが観測可能である。残りの 40 枚は裏向きにして山札として扱う。各プレイヤーは自分のターンに情報提供、カードの廃棄、カードのプレイの三つの内いずれか一つの行動を選択して

行うことができる。

情報提供では、相手プレイヤーの手札のカードの数字もしくは色を相手に教えることができる。この時、情報トークンを 1 消費する。情報トークンが 1 つもない場合、情報提供を行うことはできない。また、情報を教える際には該当するものはもれなく教えなければならない。例えば相手の手札が赤 1、赤 2、白 1、白 2、緑 1 であるときに赤色を教える場合、「1 番目と 2 番目のカードが赤である」というように教えることになり、どちらか一つのみの情報を教えることはできない。

カードの廃棄では、自分の手札のうち必要ないと考えた 1 枚を捨てることができる。その後、自分は山札からカードを 1 枚補充する。またこの時情報トークンが 7 枚以下ならば情報トークンが 1 枚増える。廃棄したカードは両プレイヤーに公開され、そのゲーム中は使用することができなくなる。

カードのプレイでは、自分の手札の内 1 枚を花火につなげることを試みることができる。プレイしたカードが盤面の同色の花火の数字より 1 大きい場合、プレイは成功となり、そのカードが盤面の該当する色の花火に並べられて数字が 1 増加する。プレイに失敗した場合、そのカードは廃棄される。成功した場合でも失敗した場合でもプレイしたカードは両プレイヤーに公開されてプレイヤーはプレイしたカードの代わりにカードを 1 枚山札から加える。

ゲームの終了条件は

- ・カードのプレイに 3 回失敗した場合
 - ・山札が 0 枚になってから各プレイヤーが行動を一回ずつ行った場合
 - ・すべての色の花火を 5 まで完成させた場合
- の 3 つである。盤面の花火の数の合計がそのまま得点になり、最高得点は 25 点である。

3. エージェントのアルゴリズム

本研究では相手プレイヤーの思考時間を戦略の指標に用いない従来の自己推定エージェントと相手プレイヤーの思考時間を戦略の指標に用いる新しい自己推定エージェントの 2 種類を Visual c++を用いて実装した。

3.1 従来の自己推定戦略

従来の自己推定戦略は我々が以前に開発したアルゴリズムであり[3]、以下の優先順位で行動を行う。

1. 数の確定情報に基づくプレイ可能カードのプレイ
色が不明だが、数字からプレイ可能カードであることが分かるカードをプレイする。例えば、ゲーム開始直後の状態の時、1 のカードは色にかかわらずプレイ可能カードになる。
2. プレイ可能カードの情報提示
相手がプレイ可能カードを持つことが分かり、そのカードの情報が相手にとって不完備な場合、情報提示を行う。数

も色も情報が不完備な場合、教える情報をランダムに選択する。

3. 自己推定に基づくプレイ可能カードのプレイ

相手の観測のシミュレーションを用いて自己の手札の推定を行う。この時考えられる手札の可能性の集合を考え、この中におけるカードの要素を出現数でソートし、1番多い出現数の値が次に多い出現数の値に対して a 倍以上のとき、このカードは1番多い出現数の値に等しいと推定する。この値がプレイ可能カードの値に等しい場合にプレイする。

4. 廃棄可能カードの情報提示

相手が廃棄可能カードを持つことが分かり、数または色を教えることによって相手視点からも廃棄可能角であると断定できる場合、その情報を提示する。

5. プレイ可能カードのプレイ

自分の手札の情報と盤面の観測可能なカードからプレイ可能カードが存在することがわかる場合、そのカードをプレイする。

6. 廃棄可能カードの廃棄

自分の手札の情報と盤面の観測可能なカードから廃棄可能カードが存在することがわかる場合、そのカードを廃棄する。

7. ランダム廃棄

自分の手札の中で、最も情報量の少ないカードを廃棄する。最も情報量の少ないカードが複数存在する場合、その中からランダムで廃棄する。

3.2 思考時間に応じて推定の信頼度を変える自己推定戦略

本研究で開発した自己推定戦略の基本的なアルゴリズムは従来の自己推定戦略と変わらない。ただし、相手プレイヤーの思考時間に応じて推定の信頼度を変更するところが従来の自己推定戦略と異なる。

まず、相手プレイヤーの各ターンにおける行動選択にかかる時間を直前の5ターン分記録する。そこから思考時間の平均と標準偏差を導き出し、次の思考時間が平均+1標準偏差以上だったら相手は長考していると定義し、思考時間が平均-1標準偏差以下だったら相手は即決していると定義する。また、直前5ターンの思考時間の偏りによって平均より標準偏差の方が大きくなり、即決であると定義できる時間がなくなってしまうことが起こりうる。そのため、相手プレイヤーが4.13秒以下の思考時間だった場合、平均や標準偏差にかかわらず即決であると定義する。この値は我々が以前大学生12人を対象として行ったHanabiの実験における各プレイヤーの平均思考時間-1標準偏差の平均をとったものである。この値は、各プレイヤーが即決したと判断される閾値の平均であるため、個人の直前の思考時間の平均や標準偏差にかかわらず即決であると判断できる閾値として妥当であると考えられる。

ここで思考時間の長短に対して以下の仮説を立てる。

- ・長考している時、相手プレイヤーは自身の行動に自信がない
- ・即決している時、相手プレイヤーは自身の行動に自信がある。

相手が行動の選択において長考を続けるとき、相手プレイヤーは選択に自信が持てていない、すなわちエージェントの戦略があまり理解できていないと考えられる。この時にエージェントが推定を行うと、通常よりも推定に失敗する確率が高くなると考えられる。なぜならエージェントの自己推定は相手が自分のシミュレーション上の相手と同じように行動を起こすことが前提となっており、そのためには相手がエージェントの戦略をある程度理解していることが不可欠だからである。そのため、このエージェントは3.1の「3.自己推定に基づくプレイ可能カードのプレイ」で示したように推定で考えられる最も多いカードの値が次に多いカードの値の a 倍以上なら推定に従うというアルゴリズムだが、相手が長考を2連続で行った場合はその a の値を大きくすることにより推定の信頼度を下げる。

一方相手が行動の選択において即決を続けるとき、相手プレイヤーは選択に自信を持っていると考えられるがその状況は大きく2つに分けられる。1つはエージェントの戦略を正しく理解できているときで、もう1つはエージェントの戦略を誤って解釈しており、その誤った解釈に基づいて行動を進めている場合である。前者の場合、AIは推定の信頼度を上げるべきだが、後者の場合は推定の信頼度を下げるべきである。そのため、この2つの状況を判別するために相手プレイヤーのカードプレイの結果を利用した。具体的には、相手プレイヤーが即決を2連続で行った場合は推定の信頼度を上げるが、相手が即決でかつカードのプレイに失敗したときには推定の信頼度を下げるようにした。

4. 評価

3.2における仮説の妥当性、また人間・AI間のゲームにおいて相手の思考時間の長短を推定の信頼度の変化の指標として用いるエージェントを用いた場合のゲームの得点に与える影響と、人間のエージェントに対して抱く印象を評価するために実験を行った。

4.1 被験者 UI

被験者とAI間のゲームを実現するために我々が開発したインタフェースを用いた。これを図1に示す。

画面左下には自分の手札、画面左上には相手の手札が表示される。中央上は現在の盤面を表し、中央下の数は左から順番に山札の枚数、情報トークンの枚数、プレイの失敗回数、得点を示す。また、右にはプレイする、プレイに失敗する、廃棄する行為によって観測可能になったカードが表示される。

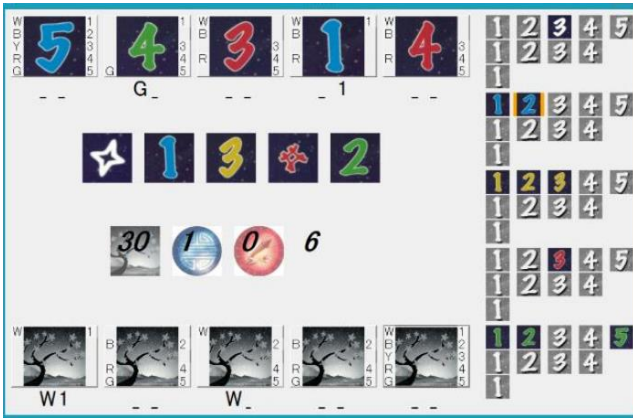


図 1 Hanabi のインタフェース

Figure 1 The interface of Hanabi

4.2 評価方法

本研究では相手のエージェントに対する印象を調べるために、ゲームの記録と被験者の様子の撮影と実験後の印象評価アンケートを行った。以下にアンケートの質問内容を示す。

- 相手についての印象
- Q1. 意思を持っていると感じたか
- Q2. ゲームに慣れていたか
- Q3. 親しみやすかったか
- Q4. 自分に配慮していたか
- Q5. 賢かったか
- Q6. 相手は自分の思い通りに行動していたか
- Q7. 相手の行動の意図が分かったか
- Q8. 相手の行動に一貫性を感じたか
- ゲームについての印象
- Q9. 得点を上げようと心掛けたか
- Q10. ゲームはやりやすかったか
- Q11. 失敗した行動があったか
- Q12. 正しいかどうか自信が持てない行動があったか
- Q13. 主体的にゲームを進めたか
- Q14. ゲームの結果に満足しているか
- Q15. 自由記述欄

Q1 から Q5 は擬人化 AI を評価する指標のひとつである Bartneck らの The Godspeed Questionnaire Series の一部を用いた[4]。Q15 の自由記述を除くすべての質問はそう思わないを 1、そう思うを 7 とした 7 段階のリッカート尺度で回答する形式を用いた。

4.3 実験手順

実験には 20 代大学生 4 人が参加した。参加者は Hanabi のルールとインタフェースについて説明を受けた後、ゲームを 2 回行った。従来の自己推定エージェントとのゲーム(以下 R 条件)、相手の思考時間に応じた推定の信頼度の変化を加えた新しい自己推定エージェントとのゲーム(以下 C 条件)の 2 条件のゲームを 1 回ずつ行った。よって R 条

件、C 条件のサンプルの数はそれぞれ 4 つになる。ゲームのプレイの際に、自分の行動選択に自信があるか否かをターンごとに逐次記録してもらった。実験に用いる山札は 2 種類用意し、1 人の被験者が 2 種類の山札で 2 回ゲームを行うようにした。ゲーム終了毎にアンケートを行い実験を終了した。

5. 結果と考察

自分の行動選択に自信があると答えたターンの平均思考時間は 21.1(SD 15.1)であるのに対して、自分の行動に自信がないと答えたターンの平均思考時間は 39.8(SD 17.8)であった。またゲーム全体を通しての平均思考時間は 24.7 s (SD17.7)であった。t 検定($p<0.05$)の結果、自信があると答えたターンの平均思考時間とゲーム全体を通しての平均思考時間($p<0.05$)及び自分の行動に自信がないと答えたターンの平均思考時間とゲーム全体を通しての平均思考時間($p<0.01$)にはどちらも有意差が見られた。このことから、思考時間が長い時プレイヤーは自身の行動選択に自信を持てていないという仮説及び思考時間が短い時プレイヤーは自身の行動選択に自信が持てているという仮説は妥当であるといえる。

R 条件における平均得点は 17.5 点(SD 5.4)であり、C 条件における平均得点は 17.8 点(SD 2.9)であった。t 検定の結果この 2 つに有意差は見られなかった。

R 条件においてカードの推定は 1 ゲームで平均 7.5 回(SD 5.7)行われ、その成功率は平均 0.57(SD 0.34)であった。C 条件においてカードの推定は 1 ゲームで平均 6.5 回(SD 2.18)行われ、その成功率は平均 0.57(SD 0.17)であった。t 検定の結果こちらも有意差は見られなかった。

以上の結果から、本研究で開発した思考時間を戦略の指標に取り入れたエージェントは得点においては影響を及ぼさないといえる。しかし、今回の研究では相手の思考時間の長短を推定の信頼度を変更する指標にしか用いていない。そのため、思考時間を、Hanabi の機械学習におけるパラメータなど、他の手段で用いることによって有意な結果が出ることを期待される。

また、アンケートの結果を表 1 に示す。表 1 より、Q1 と Q2 は R 条件と C 条件において有意差が見られた($p<0.05$)。このことから、思考時間において推定の信頼度を変更すると、推定の信頼度を変更しない場合に比べて人間は、エージェントには意思がなくゲームに慣れていないと感じることがわかる。これは、すべての実験参加者において長考によって推定の信頼度が下がる時が存在しており、その時にエージェントは確定情報をより重視して行動を行うようになり、その状態が人間側から見るとゲームに慣れておらず盤面のみをみて行動する意思のないエージェントに見えるからではないかと考えられる。

表 1 アンケートの結果

Table 1 The result of questionnaire

	R 条件平均	C 条件平均	p 値
Q1	5.50(SD 1.34)	3.50(SD 1.84)	0.016*
Q2	6.25(SD 0.43)	3.25(SD 1.60)	0.046*
Q3	3.25(SD 1.16)	2.50(SD 0.77)	0.547
Q4	3.75(SD 1.30)	3.50(SD 1.61)	0.889
Q5	3.75(SD 0.97)	2.75(SD 1.83)	0.474
Q6	3.25(SD 1.79)	2.25(SD 1.47)	0.613
Q7	6.00(SD 0.63)	3.50(SD 1.34)	0.127
Q8	5.25(SD 1.30)	4.50(SD 2.05)	0.638
Q9	6.25(SD 0.74)	5.75(SD 1.16)	0.664
Q10	5.25(SD 2.05)	3.75(SD 2.04)	0.576
Q11	5.25(SD 1.83)	6.50(SD 0.77)	0.391
Q12	5.75(SD 1.09)	6.00(SD 0.89)	0.638
Q13	5.00(SD 1.67)	4.50(SD 1.34)	0.804
Q14	4.00(SD 2.55)	2.50(SD 1.84)	0.547

*:p<0.05

6. 結論

本研究では、協力ゲーム Hanabi において、相手の思考時間から相手の行動選択の自信の有無を把握し、戦略における推定の信頼度を変更する指標として用いたエージェントを開発して、従来の自己推定戦略と比べた得点の変化、および人間に与える印象の違いを調査した。

結果として、相手の思考時間の長短と自身の有無の関係に関する仮説は妥当性が示された。また、相手の思考時間の長短に応じて推定の信頼度を変えるエージェントは、従来のエージェントと比較して得点と推定の成功率に有意差は見られなかった。さらに、推定の信頼度を相手の思考時間に応じて変更することにより、少ないサンプル数から導き出された結論ではあるが、人間はエージェントを意思がなくゲームに慣れていないエージェントであると評価することがわかった。

今回の研究では相手の思考時間を利用した戦略をとるエージェントの、従来のエージェントと比べた有用性は示されなかった。今後の課題として思考時間をより適切に戦略の指標に組み込んだエージェントの開発が考えられる。人間の思考時間に関する仮説をより複雑にして人間の本来の思考に近いものにして、思考時間の長短の定義の仕方をより適切なものに変更することによりこの課題の達成が期待される。

参考文献

- [1] Osawa, H. (2015). "Solving Hanabi : Estimating Hands by Opponent's Actions in Cooperative Game with Incomplete Information." AAAI Workshop, Computer Poker and Imperfect Information, (pp. 37–43).
- [2] Eger, M., Martens, C., & Córdoba, M. A. (2017). "An intentional AI for Hanabi" In Computational Intelligence and Games, 2017 IEEE Conference on (pp. 68–75).
- [3] 加藤 拓也, 大澤 博隆 "他者視点の推定を用いたエージェント評価の向上:カードゲーム Hanabi の AI との対戦分析から" 人工知能学会全国大会論文集, 第 32 回全国大会,4J2-01
- [4] Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots. International Journal of Social Robotics, 1(1), (pp.71–81)