1F2-3

協力ゲーム Hanabi におけるエージェント間の協調行動の分析

Estimation of Own State by Opponent's Behavior in Cooperative Game Hanabi

大澤 博隆[*1]
Hirotaka OSAWA

[*1] 筑波大学
University of Tsukuba

We used a card game called Hanabi as an evaluation task of imitating human reflective intelligence with artificial intelligence. We compared human play with random strategy, rational strategy with opponent's viewpoint, and rational strategy with feedbacks from simulated opponent's viewpoints. The results indicate that the strategy with feedbacks from simulated opponent's viewpoints achieves more score than that with the rational only strategy. The result indicates that the scores of simulation with these strategies are close to the average score of human players

1. はじめに

他者の意図を読み取る社会的な知能課題[1]の中で最も難しい課題の一つは、他者の行動から自分自身の状態を推定することである。このように他人の行動を鑑とする振る舞いは、人間の知能の特徴の一つといえる。例えば心理学分野では、自分から見えない自己像を blind self と呼び[2]、blind self の自己認識が成長の目安とされる。

本研究ではこのような blind self を競う人工知能課題として、Hanabi と呼ばれる協力ゲームを用いる。Hanabi は全エージェントが協力して得点を集める協力型のカードゲームである。全てのプレイヤは、1~5までのカードの列で表される5色の花火を協力して作り上げる。そして、この花火の大きさが得点となる。このゲームではプレイヤは自分のカードは見えないかわり、自分のカード以外の全てのエージェントのカードを知ることができる。また、このゲームではエージェント同士のコミュニケーションが制約されている。各エージェントは他のプレイヤのカードの数もしくは色を教えるため、情報伝達用の資源を消費しなければならない。その他のコミュニケーション手段は用意されない。

筆者はこの Hanabi を解くための人工知能エージェントを実装した。本エージェントは他者の視点とその行動をシミュレートできる。これによって、他者の視点の再現がどのように得点に結びつくかを検討する。

2. 関連研究

ゲームをプレイするエージェントの作成は、人工知能研究におけるランドマーク課題の一つである[3]。完全情報ゲームとして、チェッカー、オセロ、チェス、将棋、囲碁といった課題が取り組まれてきた[4][5]。これらのゲームでは、全ての情報は両方のプレイヤから観測可能であり、エージェントは勝利のために、必ずしも他者の意図を推測する必要はない。

これに対し、カードゲームなどのゲームには、他者の情報が観測不可能な不完全情報ゲームが存在する[6]。これらのゲームも研究対象となっている。ポーカーはその中でもよく知られた例であり、いくつかの理論的な分析や大会が行われている[7][8]。この他に、ブリッジや闘地主といったゲームに関する研究が知られている[9][10]。最近では、コミュニケーションゲーム人狼のように、協力者がわからない状態で行う不完全情報ゲームも課題となっている[11][12]。

これらのゲームと比較し、Hanabi はマルチエージェント課題、人工知能課題として3つの特色を持っている。まず、このゲームは協力ゲームであり、マルチエージェント間の協調を要求される課題である。また、全てのプレイヤは自分以外のプレイヤのカードを観測可能である。これは客観的な視点を持ったプレイヤが存在しないことを意味しており、リーダー不在の状況での協調を求められる課題である。最後に、Hanabi ではプレイヤ間のコミュニケーションが厳しく制約されている。プレイヤは他のプレイヤの色、もしくは数字を教えることができるだけであり、その教示には情報カウンタと呼ばれる資源を消費する必要がある。このように制限された条件のため、Hanabi の解法では自然言語処理からの意味理解を必要としない。また、ゲーム理論におけるチープトークと呼ばれる利得を伴わない情報交換がない[13]。Hanabi を人工知能課題として用いることで、言語に依存しない形の相手のモデル化がどのように行われるか、といった知能の働きを検討することが可能である。

3. Hanabi のモデル化

3.1 ゲームルール

Hanabi は 2 から 5 人のプレイヤによって行われるゲームである。本研究では、2 人のプレイヤによるゲームのみを扱う。Hanabi は 5 色(白、赤、青、黄、緑)、50 枚のカードを使用する。1 色につき、10 枚のカードが存在する。1 のカードは 3 枚、2 から 4 までのカードは 2 枚、5 のカードは 1 枚存在する。本ゲームのゴールは、1 から 5 までの数字のカードが重なった、5 つの異なる色の山(花火)を作ることである。

ゲーム開始時に、各プレイヤは 5 枚のカードを配られる。残りの 40 枚のカードは山札として積まれる。また、2 人のプレイヤは 8 枚の情報カウンタを共有する。2 人のプレイヤが交互にターンを重ね、協力して花火を作成する。

各プレイヤは、ゲームの各ターンにおいて、カードの情報提供、破棄、プレイの 3 つの行動が許されている。1 つめの行動は情報提示であり、この行動を選んだプレイヤは相手の持っているカードの色か数字を教えることができる。情報を教える場合には、情報カウンタを一つ消費する。例えば、相手のカードが赤 1、緑 2、緑 3、白 2、白 4 である場合、相手に緑の色を教える場合には、2 番目のカードと 3 番目のカードが緑である、と教えることが可能である。また、数字 2 を教える場合には、2 番目と 4 番目

連絡先:大澤博隆, システム情報系知能機能工学域, 〒305-8573 茨城県つくば市天王台 1-1-1 , osawa@iit.tsukuba.ac.jp

のカードが 2 であると教え、どちらか一方しか教えない、ということはできない。情報カウンタが存在しない場合には、相手に情報を教えることはできない。

2 つめの行動は、カードの破棄である。プレイヤは自分が持っている 5 枚のカードのうち、必要ないと考えるカードを捨て、その代わりに新しいカードを山札から補充し、さらに共有する情報カウンタを戻すことができる。破棄したカードは自分を含めた全員に公開される。一度破棄したカードは、2 度と使うことはできない。例えば、1 の数字のカードは同じ色のものが 3 枚あるため、どれかを捨てても、他 2 枚のどちらかのカードを使うことで、花火を完成させることができる。しかしながら、5 のカードは各色 1 枚ずつしかないため、このカードを捨ててしまうと、この色の花火が完成することはない。また、情報カウンタが既に 8 個ある場合には、カードを捨てることはできない。

3 つめの行動は、カードのプレイである。プレイヤは自分が持つカードのうち、どれか一つを花火につなげる「プレイ」を行うことができる。もし、自分の出したカードが既存の同色の花火よりも 1 つだけ大きい数字の場合、花火を成長させることができる。例えば、緑の花火が 1 から 3 までの数字で構成されている時、緑 4 のカードを出すことで、緑の花火を 1 から 4 までの数字に成長させることができる。ただし、緑 3 や緑 5 のように、繋がらない数字を出してしまうと、それは失敗となり、出したカードは花火に接続されず破棄され、ゲーム中で二度と使用できなくなる。失敗が 3 回繰り返されると、ゲームはその時点で終了する。また、その色の花火が存在しない時にその色の 1 のカードを出すと、新しく花火を作ることが可能である。プレイ後に、カードを補充するためプレイヤは山札から新たにカードを 1 枚加える。なおオリジナルのゲームでは、ある色の花火が 1 から 5 までのカードを揃えた場合、成功となり情報カウンタを 1 つ戻すことができる、というボーナスがあるが、単純化のため今回はこのルールを採用しない。

3 回プレイが失敗する場合、山札からカードが尽きて、さらに一周した場合、全ての色の Hanabi が完成する場合の、どれか 3 つの場合にゲームが終了する。終了後には、各色の花火のそれぞれのカード枚数を合計し、これが特典となる。最大で各色 5 種×5 枚のカードで、25 点となる。

3.2 Hanabi のモデル化

以下、戦略の記述を用意にするため、Hanabi のモデル化を行う。それぞれのカードを各色の最初の文字と数字で表す。例えば、赤の 4 のカードは R4、緑の 1 のカードは G1 と表す。情報の足りない場合には、そのカードをアンダーラインの形で表し、可能なカードの集合の表記とする。例えば、R_というカードはそのカードが{R1,R2,R3,R4,R5}のどれかであることを表す。

ゲームの山札を集合 P、捨てられたカードを集合 T、各花火を集合 D で表す。P は両プレイヤから観測不可能なカードの順列集合であり、P={Y3,W1,R2,…}のように表す。T は捨てられたカードの集合であり、両プレイヤから観測可能なカードの集合となる(T={R3,G2,Y1,…})。D は各花火の集合を表しており、空集合を含めた順序集合として D = {{},{B1},{},{R1,R2},{G1,G2,G3}}のように表される。ゲームの開始時に P は 50 のカードの集合であり、T は 0、D は 5 つの空集合の集合{{},{},{},{},{}}となる。

全てのプレイヤは各プレイヤ自身の視点を保持する。これを W_{pl} や W_{co} のように表す(pl は player であり、co は cooperator(協力者)を表す)。W は 5 枚のカードの状態の順序集合 C を保持している。各プレイヤの視点上の自分のカードや相手のカードを記述可能である(C_{pl}={R_,R_,__,__,_3}, Cco={R1,R2,W1,W2,G3}など)。各プレイヤは関数 F で表される。F は入力として D, P, T, W を受け取り、出力として A を返す(A = F_{pl}(D, P, T, W_{pl}))。

もし相手がカードに関する情報を与えた場合、そのカードに対する情報は狭まる。例えば、あるカードを指して他方のプレイヤが赤と指摘した場合、そのカードが赤であると同時に、他のカードは赤でないという情報が入る。このため、可能集合は常に減少する。

もし、あるカードがプレイするのが可能であるという確定的な情報を持つ場合、そのカードを「プレイ可能カード」と定義する。例えば、場に花火が一つも出ておらず、あるカードが 1 という情報が決定している場合、そのカードは何色であるかに関わらず、プレイ可能なカードとなる。また、もし緑の花火が 3 まで完成していれば、G4 のカードはプレイ可能カードである。もし、あるカードかこれ以上プレイ可能とならないという確定的な情報を持つ場合、そのカードを破棄可能カードと定義する。例えば、場に赤4 の花火が完成している場合、R1,R2,R3 のカードはいずれも破棄可能カードである。また、黄色の花火が 5 まで完成している場合、全ての黄色のカードは、数字の情報があるなしに関わらず、破棄可能カードとなる。

あるカードと同じものが存在し、それが公開されていない場合、このカードは重複カードと定義される。例えば、R2 というカードが 1 枚存在し、それ以外のカードが公開されていなければ、R2 は重複カードと定義される。

4. Hanabi の戦略

4.1 完全戦略

Hanabi の各プレイヤが最高得点を取る場合は、お互いに情報を得ている場合である。この場合には、両方のプレイヤは常に最適な手を打つことができる(図 1 左上)。Hanabi のルール上、両プレイヤから見えない山札などの情報にアクセス出来ない限り、この戦略に勝てる戦略は存在しない。得点の比較のため、最高得点を取るような戦略を持つプログラム作成した。この戦略は、以下の様な手順で推移する。

1. もしプレイヤがプレイ可能カードを持っていれば、そのカードをプレイする。
2. もし相手がプレイ可能カードを持っていれば、プレイヤは情報を教えて相手に番を回す(両者が情報を保持している場合、情報を教える意味は無いため、実質的にターンスキップと同じである)
3. もしプレイ可能なカードが両者に無い場合は
   - もしプレイヤが破棄可能カードを持っていたら、それを捨てる
   - もしプレイヤが破棄可能カードを持っておらず、重複カードを持っていた場合、それを捨てる
   - もしプレイヤが破棄可能カードも重複カードも持っていない場合には、手の中から一番大きい数字のカードを捨てる。

4.2 ランダム戦略

ランダム戦略では、プレイヤはカードに対するあらゆる情報を持たない(図 1 左中央)。このような戦略はもっとも得点の低い戦略であると想定できる。本ランダム戦略では、30%の確率で情報

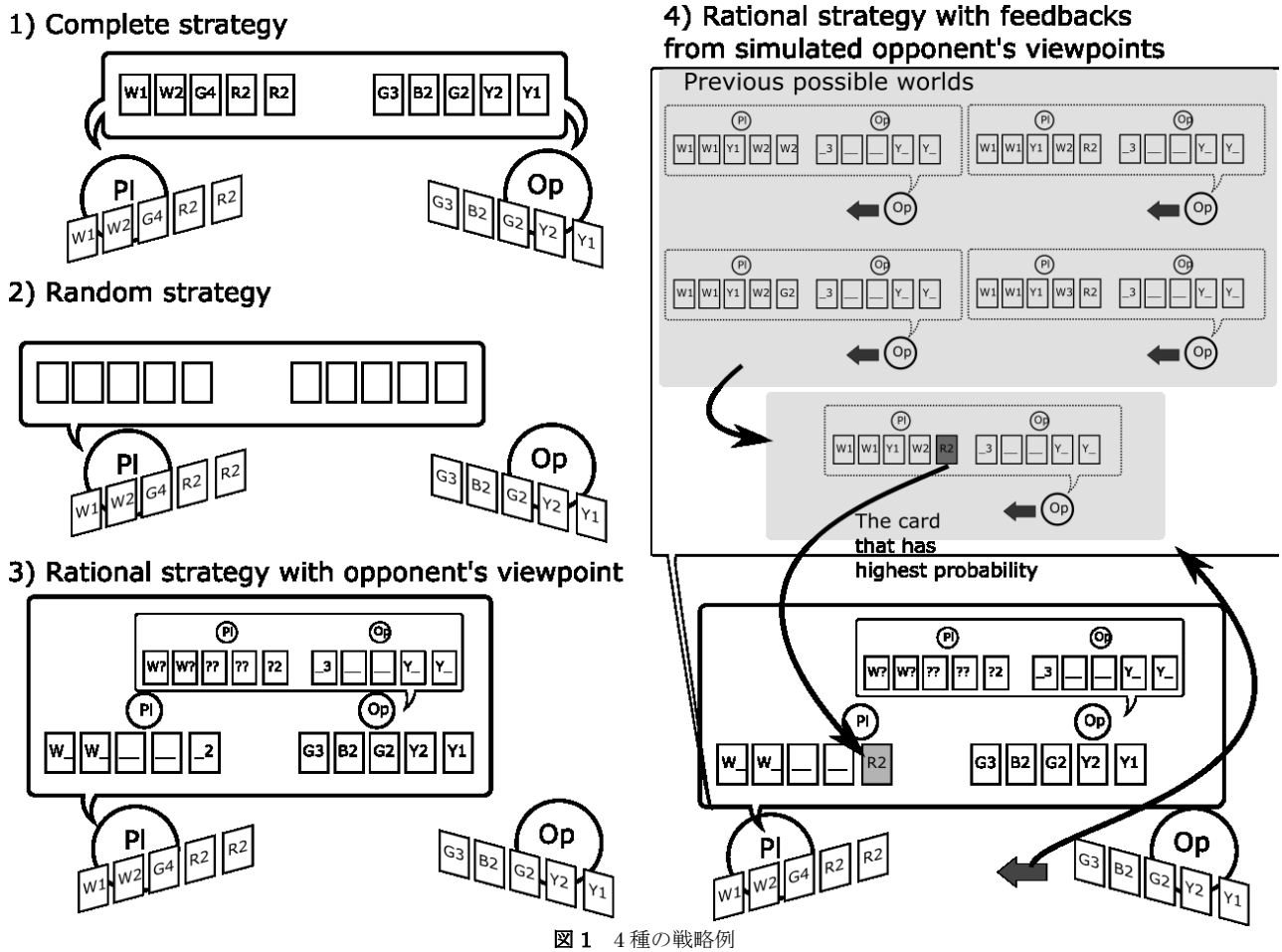

図 1　4 種の戦略例

提示、40%の確率でランダムなカードの破棄、30%の確率でランダムなカードのプレイを行った。

4.3 決定論的戦略

合理的なプレイヤは、自分の教えられた手と相手の手を覚えることができる(図 1 左下)。これは確定的な情報であり、こうした確定的な情報下でプレイを行うエージェントは、確率的な行動を除けばもっとも合理的なプレイが可能となる。具体的な手順を下記に示す。。

1. もしプレイヤがプレイ可能カードを持っていれば、そのカードをプレイする。
2. もしプレイヤが破棄可能カードを持っていたら、それを捨てる
3. もし相手がプレイ可能カードを持っていれば、プレイヤはそのカードの色か数字情報をどれか一つ教える
4. もし相手がプレイ可能カードを持っていない場合、プレイヤは相手が情報を持っていないカードのうち1枚をランダムに選び、そのカードの色か数字の情報を教える
5. もし相手がプレイ可能なカードを持っておらず、プレイヤに情報トークンが存在しない場合、プレイヤは自分のカードを1枚ランダムに選び、それを捨てる。

4.4 他者行動からの自己推定戦略

本戦略では、他者視点のある合理的戦略と同じようにプレイを行うが、4.3 のステップ 4 で、相手がプレイ可能なカードをもっていない場合、相手の出した手から、相手の視点をシミュレートし、その結果として自分のカードを予想する(図 1 右)。これは相手が自分と同じように考える、という前提を元にした確率的推論である。この予想の手順は以下のとおりである。

プレイヤは自分が持つ手の可能な組み合わせ集合 H を全て考える(例えば、H={{R1, R1, G2, G2, W1}, {R1, R1, G2, G2, W2}…})プレイヤは、一手前のゲームの状態 D_{pre}, P_{pre}, T_{pre}を再現する。そして、H の一つ一つの要素を P_{pre}に当てはめる。当てはめた P_{pre} の各要素に対して、シミュレートした F_{op} の結果を計算する(A_{hyp}=Fop (D_{pre}, P_{pre}, T_{pre}, W_{hyp_op}))もしシミュレートした結果 A_{hyp}が、相手が一手前に行った現実の手と一致しない場合、その手を集合 H から取り除く。この操作を H から取り除ける要素がなくなるまで繰り返す。以上の手続きにより、プレイヤは自分の手の可能集合 $H_{estimate}$を求めることができる。

プレイヤはヒューリスティクスを用いて持っているカードを推測する。集合のうち、最も登場したカードの総数を x、次に登場したカードの総数を y とする。x が y の a 倍より大きい場合に、そのカードを持っていると考える。例えば、花火がもしひとつも完成していないときに、相手のプレイヤが「あなたの一番右のカードは緑」と教えるとする。このカードに対し情報が与えられる可能性は、プレイ可能カードであるときに大きくなる。従って計算により、このカードを緑、と教えるのは、このカードが G1 であるとき、という可能性が多くなる、以上の手続きより、このカードを G1 と推測し、1 が出せるという合理的規則に基づいて、プレイヤはこのカードをプレイする。

5. シミュレーション

5 つの戦略について、それぞれコンピュータで 100 回シミュレートした結果と、人間のプレイヤで 20 回シミュレートした結果を比較した。条件はいずれも同条件である。計算では、持てるカードが 2 枚の場合、5 枚の場合のそれぞれの計算を行った。資源の制限のため、戦略 5 の再帰的なシミュレーション推定は直前の一回のみ行った。事前のシミュレーションにより、a=2.5 という値が戦略 5 において最も高得点であったため、この値を採用した。

結果として、完全戦略の際の得点は 24.6 (SD 1.10) 、ランダムの際は 2.20 (SD 1.60) 、他者視点のある合理的戦略では 14.53 (SD 2.24) 、他者視点のシミュレートをフィードバックする合理的戦略では 15.85 (SD 2.26) となった。ANOVA 検定を行った結果、全ての対において、$p<.05$ となり、全群の有意差が示された。本結果は、他者視点のシミュレートをフィードバックする合理的戦略が、単なる合理的な戦略以上に良い結果を示したことを示唆する。

6. 考察と結論

本結果は、他者の視点から自己の視点を推測するシミュレーションが、ゲームの得点上昇に対して有意に働いていることを示唆する。たとえば、D={W:5, R:5, B:0, Y:3, G:2} 、W_{pre_pl}={C_{pre_pl}=(__,_1,__,Y_,__)、C_{pre_op}=(Y1, R4, Y4, B2, Y4)}、かつ W_{pre_op}={C_{pre_pl}=(Y3, B1, G1, Y2, B1),、C_{pre_op}=(__, __, _4, __, _4)}というとき、プレイヤが相手に対して 1 のカードを教えた場合があった(Cop=(_1,__,_4,__,_4))。このとき、教えられたカードは Y1 である可能性がもっとも高いとこのプログラムは判断し、Y1 をプレイしている。同様の状況で、合理的戦略では手を決定できず、このような決断ができなかった。

本研究は、相手のシミュレーションによる自己推定による協力課題を解いている、と捉えることができる。筆者は現在、人狼を解く人工知能の作成をしているが、こうした相手のモデルを想定する知能は人狼課題を解く上でも有用であると考えられる。

7. 謝辞

本研究は JSPS 科研費 25700024 の助成を受けたものです。

参考文献

[1] R. W. Byrne and A. Whiten, *Machiavellian Intelligence: Social Expertise and the Evolution of Intellect in Monkeys, Apes, and Humans*. Oxford University Press, USA, 1989.

[2] J. Luft and H. Ingham, “The Johari Window: a graphic model of awareness in interpersonal relations,” *Hum. relations Train. news*, vol. 5, no. 9, pp. 6–7, 1961.

[3] B. Abramson, “Control strategies for two-player games,” *ACM Comput. Surv.*, vol. 21, no. 2, pp. 137–161, Jun. 1989.

[4] K. Krawiec and M. G. Szubert, “Learning n-tuple networks for othello by coevolutionary gradient search,” in *Proceedings of the 13th annual conference on Genetic and evolutionary computation - GECCO ’11*, 2011, pp. 355–362.

[5] S. Gelly, L. Kocsis, M. Schoenauer, M. Sebag, D. Silver, C. Szepesvári, and O. Teytaud, “The grand challenge of computer Go,” *Commun. ACM*, vol. 55, no. 3, p. 106, Mar. 2012.

[6] S. Ganzfried and T. Sandholm, “Game theory-based opponent modeling in large imperfect-information games,” in *International Conference on Autonomous Agents and Multiagent Systems*, 2011, pp. 533–540.

[7] D. Billings, D. Papp, J. Schaeffer, and D. Szafron, “Opponent Modeling in Poker,” in *AAAI Conference on Artificial Intelligence*, 1998, pp. 493–499.

[8] D. Billings, N. Burch, A. Davidson, R. Holte, J. Schaeffer, T. Schauenberg, and D. Szafron, “Approximating Game-Theoretic Optimal Strategies for Full-scale Poker,” in *International Joint Conference on Artificial Intelligence*, 2003, pp. 661–668.

[9] M. L. Ginsberg, “GIB: Imperfect Information in a Computationally Challenging Game,” *J. Artif. Intell. Res.*, vol. 14, pp. 303–358, 2001.

[10] D. Whitehouse, E. J. Powley, and P. I. Cowling, “Determinization and information set Monte Carlo Tree Search for the card game Dou Di Zhu,” in *2011 IEEE Conference on Computational Intelligence and Games (CIG’11)*, 2011, pp. 87–94.

[11] **片上大輔**, **鳥海不二夫**, **大澤博隆**, **稲葉通将**, **篠田孝祐**, and **松原仁**, “**人狼知能プロジェクト**,” **人工知能**, vol. 30, no. 1, pp. 65–73, 2015.

[12] R. Aylett, L. Hall, S. Tazzyman, B. Endrass, E. André, C. Ritter, A. Nazir, A. Paiva, G. Höfstede, and A. Kappas, “Werewolves, cheats, and cultural sensitivity,” in *Autonomous Agents and Multi-Agent Systems*, 2014, pp. 1085–1092.

[13] K. Wärneryd, “Evolutionary stability in unanimity games with cheap talk,” *Econ. Lett.*, vol. 36, no. 4, pp. 375–378, Aug. 1991.