

MCCFR アルゴリズムによる 5 人狼ゲームの戦略の学習

小島 大征^{1,a)} 伊藤 暢浩¹ 岩田 員典² 内種 岳詞¹

概要：近年，多人数不完全情報ゲームの研究に注目が集まっている．本研究では，多人数不完全情報ゲームである 5 人狼ゲームに対して，他の不完全情報ゲームで有効とされている Monte Carlo CFR アルゴリズムを適用し，その有効性について検討する．MCCFR を適用するために簡易化した 5 人狼ゲームを構築し戦略を学習する．有効性の評価のために多様なエージェントとの対戦をおこなった結果，勝率を向上させる戦略を学習できることを示した．

キーワード：多人数不完全情報ゲーム，Monte Carlo CFR，人狼ゲーム

Learning Strategy by MCCFR Algorithm for the 5-Player Werewolf Game

Abstract: In recent years, the study of multi-player imperfect information games has attracted much attention. In this paper, we examine the effectiveness of the Monte Carlo CFR (MCCFR) algorithm, a method for calculating strategies of imperfect information games, for the 5-player werewolf game. In order to apply MCCFR, we constructed the simplified 5-player werewolf game and learning process by MCCFR. We played against various agents to evaluate the effectiveness of the learned strategy by MCCFR. As a result, the method improved the win rate of the 5-player werewolf game.

Keywords: Multi-player imperfect information game, Monte Carlo CFR, Werewolf game

1. はじめに

近年，不完全情報ゲームの研究が盛んにおこなわれている．2 人零和不完全情報ゲームでは理論的に最適な戦略を求めるための手法が提案されており，また人間のプロのプレイヤーに勝利する AI が開発されるなど大きな成果を上げている．一方で，多人数不完全情報ゲームにおいては最適な戦略を求めるための手法は確立されていない．

多人数不完全情報ゲームの研究として，人狼ゲームが注目されている．その研究プロジェクトとして，人狼知能プロジェクトがある [1]．このプロジェクトでは人狼ゲームを計算機上でプレイするための人狼知能プラットフォームの提供や，エージェントの開発促進のために勝率を競う人

狼知能大会を開催している．

2 人零和不完全情報ゲームの研究として，2 人プレイヤーの heads-up no-limit Texas Hold’em (以降，HUNL) を対象としたゲーム AI の研究がある．HUNL では，Moravčík ら [2] による DeepStack が人間のプロプレイヤーに勝利している．また，Brown ら [3] による Libratus は，同ゲームにおいて 4 名のプロのプレイヤーに勝利している．DeepStack や Libratus は，2 人零和不完全情報ゲームにおいてナッシュ均衡戦略を求めるための代表的な手法である Counterfactual Regret Minimization (以降，CFR) [4] ベースのアルゴリズムを利用して開発されている．

CFR は多人数におけるゲームでは最適な戦略に収束する保証はないが，実験的に強い戦略を得られることが報告されている [5]．Brown ら [6] は 6 人プレイヤーの HUNL において，人間のプロに勝利した AI である Pluribus を開発した．Pluribus は CFR を改良したアルゴリズムである Monte Carlo CFR (以降，MCCFR) [7] を利用して開発さ

¹ 愛知工業大学大学院
Aichi Institute of Technology

² 愛知大学
Aichi University

^{a)} k16039kk@maslab.aitech.ac.jp

れている．このように，2人零和不完全情報ゲームで成功している手法を多人数ゲームにおいて利用している事例がある．

人狼ゲームにおける関連研究では，畢ら [8] が CFR ベースのアルゴリズムを用いて戦略を求めている．畢らは，多人数でおこなわれる人狼ゲームを 2人零和不完全情報ゲームとしてモデル化して戦略を求めた．しかし，この研究では，人狼ゲームのゲーム状態数や行動数を大きく制限したゲームの上で戦略の学習をおこなっており，実際の人狼ゲームとは大きく異なっている．

以上のことから，本研究では多人数不完全情報ゲームである 5 人狼ゲームに対して，他の不完全情報ゲームで有効とされている MCCFR [7] を適用し，その有効性について検討する．MCCFR の適用のために簡易化した 5 人狼ゲームの構築をおこない，戦略を学習する．有効性の評価のために，得られた戦略に従うエージェントを作成して多様なエージェントとの対戦をおこなう．

2. 人狼ゲームと人狼知能プロジェクト

2.1 人狼ゲームの基本ルール

人狼ゲームとはプレイヤーが村人陣営と人狼陣営に分かれ，対話を通して相手陣営の排除を目指すコミュニケーションゲームである．ゲーム開始時にプレイヤーに役職が割当てられ，各役職に応じた能力を使用して他陣営の排除を目指す．

人狼ゲームは，1 日が昼と夜のフェーズに分かれており 1 日を単位として進行する．昼のフェーズは，誰が人狼であるかについて議論し，投票による多数決で 1 名をゲームから排除する．夜のフェーズは，人狼役職による襲撃により 1 名をゲームから排除する．また，役職によって与えられた能力を行使できる．

村人陣営の勝利条件はゲームからすべての人狼役職を投票により排除することである．反対に人狼陣営の勝利条件は，ゲームに残っている人狼役職の人数が人間の数と同数以上となることである．ゲームは，どちらかの陣営が勝利するまでおこなわれる．

2.2 人狼知能プロジェクト

人狼知能プロジェクトは，人狼ゲームをプレイする人工知能の構築を目指すプロジェクトである [1]．同プロジェクトでは，人狼ゲームをプレイする人工知能のことを「人狼知能エージェント（以降，エージェント）」と呼ぶ．

同プロジェクトはこの目的を実現するために，計算機上で人狼ゲームをプレイするための人狼知能プラットフォームの提供をおこなっている．また，エージェントの勝率を競う人狼知能大会を開催している．

2.3 人狼知能プラットフォーム

人狼知能プラットフォームとは，計算機上で人狼ゲーム

表 1 5 人狼ゲームにおける役職と配役数

Table 1 Roles in the 5-player werewolf game

役職名	配役数	陣営	種族	特殊能力
村人	2	村人	人間	なし
占い師	1	村人	人間	任意のプレイヤーの種族を知ることができる
人狼	1	人狼	人狼	任意のプレイヤーを襲撃できる
裏切り者	1	人狼	人間	なし

をおこなうためのプラットフォームである．また，同プラットフォームではエージェントの開発環境も備えている．

人狼知能プロジェクトは，人狼ゲームを計算機上でプレイするために行動や発言を抽象化した人狼知能プロトコルを定義している [1]．人狼知能プラットフォームは，このプロトコルを用いて人狼ゲームをおこなう．

人狼知能プラットフォームでは，ゲーム開始時に役職と ID が割当てられる．各エージェントは，この ID により互いを識別する．人狼知能プラットフォームでは，0 日目は夜のフェーズのみおこなわれるが，ここでは占い師役職という役職に限って能力が使用できる．そして，昼のフェーズにおける議論はターン制でおこなわれ，エージェントの発言は同時に発言されたものとして扱われる．エージェントは，各ターンに 1 回発言することができる．また，発言は最大で 10 ターン分おこなえるが，発言しないこともできる．議論は 20 ターンが経過したか，または誰も発言しないターンが一定数経過したら終了する．なお，2020 年からの人狼知能大会で用いられている人狼知能プラットフォームでは，ターン内で発言する順番がランダムに与えられて，同ターン内の場合でも自分より前の発言を受け取れる仕様となったが，本研究では，2020 年より前の仕様に基づいて研究をおこなった．

2.4 人狼知能大会

人狼知能大会では，5 体のエージェントによる 5 人狼ゲームと，15 体のエージェントによる 15 人狼ゲームで対戦がおこなわれる．100 ゲームを 1 セットとして，時間の許す限りゲームの実行がおこなわれ，これらのゲームの合計勝率により順位付けされる．

2.5 5 人狼ゲーム

人狼ゲームは多人数でおこなわれるゲームだが，考慮する人数が多いほど戦略を学習することが困難になると考えられる．そのため，本研究では人狼知能大会における最少人数のゲームである 5 人狼ゲームを対象とする．

5 人狼ゲームにおける役職と配役数について表 1 に示す．5 人狼ゲームの役職の配役数は村人 2 人，占い師 1 人，裏切り者 1 人，人狼 1 人となっている．

5 人狼ゲームでは，1 日に投票と襲撃によりゲームから 2 人排除される．そのため，ゲームは最長 2 日で終了する．人狼役職は 1 人であるため，人狼役職を投票によりゲーム

から排除できたら村人陣営の勝利となる。1日目で人狼役職のプレイヤーを排除できなかった場合は、3人のプレイヤーで2日目が始まる。2日目で人狼役職のプレイヤーを排除できなかった場合は人狼陣営の勝利となる。

3. 本研究の目的と手法

3.1 本研究の目的

本研究では、MCCFRの適用のために簡易化した5人狼ゲームを構築する。これを対象にMCCFRにより戦略を学習する。そして、学習した戦略を用いて複数のエージェントと対戦することで、MCCFRの有効性を確認する。

すでに、先行研究[9]で我々は5人狼ゲームに対してMCCFRの適用をおこなっている。先行研究では、学習で得た戦略を単一の対戦組合せのみを用いて評価した。本研究では、さらに多人数ゲームの結果が相手の戦略に依存することを考慮し、複数の対戦組合せにより戦略を評価する。また、元の人狼ゲームに近づくよう、本研究では先行研究のものから簡易化の方法を一部変更する。

3.2 展開型ゲームと戦略

MCCFR[7]やCFR[4]では、展開型ゲームを対象に戦略の学習をおこなう。展開型ゲームとは、木構造で表される不完全情報ゲームを記述するためモデルである。

有限展開型ゲームとは、プレイヤーの有限集合 N 、行動の履歴の有限集合 H 、終端履歴の集合 Z ($Z \subseteq H$)、行動の集合 A 、ある履歴 $h \in H$ に対してその履歴の後に行動するプレイヤーを表す関数 P 、チャンスプレイヤー（後述）における行動の確率分布を表す関数 f_c 、各プレイヤー $i \in N$ における利得を表す関数 u_i 、各プレイヤー $i \in N$ における情報分割 I_i からなる[4], [7]。

ゲーム状態を表す非終端履歴 h において可能な行動の集合を $A(h)$ と表し、プレイヤーは行動 $a \in A(h)$ を1つ選択する。非終端履歴 h のあとに行動するプレイヤーを表す関数を $P(h) \in N \cup \{c\}$ と表現する。このとき、 c はチャンスプレイヤーと呼ばれており $P(h) = c$ であれば、非終端履歴 h の後に取られる行動はプレイヤーの意思とは関係なく事前決められている確率分布 $f_c(h)$ にしたがって決定される。終端履歴 z に到達したら、プレイヤー i は利得 $u_i(z)$ を得る。

不完全情報ゲームではゲームの状態が部分観測であるため、実際には異なるゲーム状態であってもプレイヤーには区別できないことがある。プレイヤー i にとって区別はつかないが実際には異なるゲーム状態の集合を情報集合 $I_i \in \mathcal{I}_i$ と呼ぶ。

プレイヤー i の戦略 σ_i とは、各情報集合 I_i において可能な行動である $a \in A(I_i)$ の確率分布を返す関数である。各プレイヤーの戦略の集合 $\sigma = \{\sigma_1, \sigma_2, \dots, \sigma_N\}$ を戦略プロファイルといい、戦略プロファイルからプレイヤー i の戦略を除いたものを σ_{-i} と表す。ここで、各プレイヤーが σ に従って

行動したときに履歴 h に到達する確率を $\pi^\sigma(h)$ と表す。

3.3 Counterfactual Regret Minimization

MCCFR[7]は、Counterfactual Regret Minimization (CFR) [4], [10]を改良した手法である。CFRは展開型ゲームとしてモデル化される不完全情報ゲームにおいてナッシュ均衡となる戦略を求めるための代表的な手法である。

非終端履歴 h から戦略 σ に従った場合に counterfactual value について式(1)と定義する。

$$v_i(\sigma, h) = \sum_{z \in Z, h \sqsubset z} \pi_{-i}^\sigma(h) \pi^\sigma(h, z) u_i(z) \quad (1)$$

ここで、 $h \sqsubset z$ は h が z の子孫であることを表し、 $\pi_{-i}^\sigma(h)$ はプレイヤー i を除いたすべてのプレイヤーが戦略 σ にしたがって行動したときに履歴 h へ到達する確率、 $\pi^\sigma(h, z)$ は戦略 σ により履歴 h から終端履歴 z に到達する確率である。

また、履歴 h において、行動 a をとらないことに対する counterfactual regret を式(2)と定義する。ここで $\sigma_{I \rightarrow a}$ は、プレイヤーは戦略 σ にしたがうが情報集合 I において必ず行動 a をおこなう戦略である。

$$r_i(h, a) = v_i(\sigma_{I \rightarrow a}, h) - v_i(\sigma, h) \quad (2)$$

そして、ある情報集合 I における、行動 a をとらないことに対する counterfactual regret を式(3)と定義する。

$$r_i(I, a) = \sum_{h \in I} r_i(h, a) \quad (3)$$

CFRの反復回数 T における cumulative counterfactual regret を式(4)に定義する。

$$R_i^T(I, a) = \sum_{t=1}^T r_i^t(I, a) \quad (4)$$

$(x)^+ = \max(0, x)$ としたときに、CFRの反復回数が T であるときの戦略は cumulative counterfactual regret をもとに式(5)のように更新される。

$$\sigma_i^T(a|I) = \begin{cases} \frac{(R_i^{T-1}(I, a))^+}{\sum_{a' \in A(I)} (R_i^{T-1}(I, a'))^+} & \text{if } \sum_{a' \in A(I)} (R_i^{T-1}(I, a'))^+ > 0 \\ \frac{1}{|A(I)|} & \text{otherwise.} \end{cases} \quad (5)$$

2人零和不完全情報ゲームのとき、反復回数 T までを平均した戦略がナッシュ均衡戦略に収束することが知られている。しかし、多人数ゲームの場合では理論的に保証されていない。ただし、実験的に強い戦略を得られることが報告されている[5]。

3.4 Monte Carlo CFR (MCCFR)

CFRはゲーム木を全探索するようなアルゴリズムである。そのため、大規模なゲームにおけるゲーム木では探索が困難となる。Monte Carlo CFR (MCCFR) [7]は、1回の反復におけるゲーム木の探索部分を制限することでその

問題に対処できるアルゴリズムである。

Monte Carlo CFR (MCCFR) では、終端履歴の集合である Z の部分集合の集合 $Q = \{Q_1, Q_2, \dots, Q_r\}$ を定義する。これらの部分集合をブロックと呼ぶ。MCCFR では、各反復でブロックの1つをサンプリングし、そのブロックの終端履歴のみ考慮する。ブロック Q_j をサンプリングする場合のプレイヤー i の sampled counterfactual value を式 (6) に示す。ここで $q(z)$ とは、現在の反復において終端履歴 z を考慮する確率である。また、情報集合 I から到達できる $z \in Z$ の集合を Z_I と書き、その要素 $z \in Z_I$ について情報集合 I との共通部分を $z[I]$ と表す。

$$\tilde{v}_i(\sigma, I|j) = \sum_{z \in Q_j \cap Z_I} \frac{1}{q(z)} u_i(z) \pi_{-i}^\sigma(z[I]) \pi^\sigma(z[I], z) \quad (6)$$

sampled counterfactual value は、期待値で counterfactual value と一致するように設計されている。sampled counterfactual value を用いて、sampled counterfactual regret を、式 (7) として定義し、この値を用いて戦略の更新をおこなう。

$$\tilde{r}_i(I, a) = \tilde{v}_i(\sigma_{I \rightarrow a}^t, I) - \tilde{v}_i(\sigma^t, I) \quad (7)$$

MCCFR はいくつかのサンプリング方法が提案されている。本研究では、その中の1つである External-Sampling [7] を用いる。

3.5 MCCFR による戦略獲得について

本研究では、5 人狼ゲームに対して MCCFR アルゴリズムの適用をおこない戦略を学習する。MCCFR ではすべてのゲーム状態の regret を保持する必要がある。そのため、状態数が多いゲームの場合は学習が困難になるという問題がある。

人狼知能プロトコルでは、発言に関するコマンドが 10 種類以上ある。5 人狼において、1 日目の議論時におけるターン数を 2020 年大会のログによる 1 日目の議論時の最小ターンであった 4 として概算すると 10^{20} 以上のゲーム展開が考えられる。CFR では、 10^{12} のゲーム状態を持つポーカーを解くことが報告されている [4] が、その状態数と比べると人狼ゲームの状態数は非常に多く、学習が困難になると考えられる。そのため、人狼知能大会や人狼ゲームの定石などからゲーム状態数の制限をおこない、簡易化した 5 人狼ゲームの構築し、その上で戦略を学習する。

本研究では、簡易化した 5 人狼ゲームに対して MCCFR の適用し戦略を学習する。簡易化した 5 人狼ゲームにおけるゲーム状態は、プレイヤーが観測できる全プレイヤーの行動の履歴として作成する。ゲーム状態の例を図 1 に示す。

人狼ゲームでは、ゲーム状態によって議論時における発言の選択、投票対象の選択、特殊能力の対象を選択する必要がある。したがって、MCCFR により各ゲーム状態にお

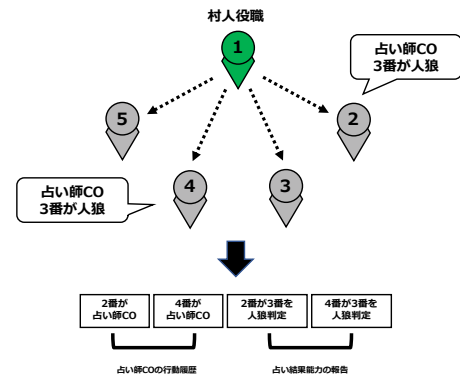


図 1 簡易化した 5 人狼ゲームにおけるゲーム状態の例

Fig. 1 Example of the game state in the simplified 5-player werewolf game

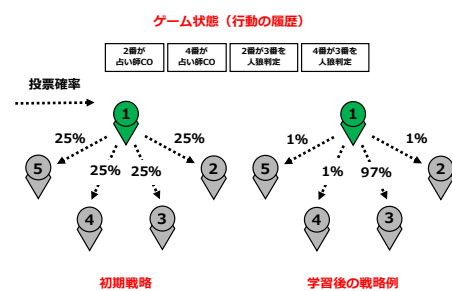


図 2 あるゲーム状態における戦略の変化の例

Fig. 2 Example of strategy changes in a game state

いてどのように行動選択すべきかについて学習する。図 2 より MCCFR による学習によって戦略の更新について、1 日目のある状態における村人役職の投票の戦略を例として説明する。図 2 より学習をおこなう前の初期戦略は、行動を等確率で選択するものとして表現できる。この図では、自身が占い師であると宣言しているプレイヤーの発言から、3 番に投票することによって勝利できる可能性が高く、投票確率が大きくなるように変化したといえる。MCCFR では各ゲーム状態に対して regret により戦略の更新がおこなわれるため、初期戦略と比べてゲームに勝利するための戦略を得ることができると考えられる。

3.6 簡易化した 5 人狼ゲーム

先行研究 [9] から一部変更をおこなった簡易化した 5 人狼ゲームを構築した。本研究では議論時における発言の種類の制限について、より人狼ゲームに近くなるように人狼知能大会のログに基づいた更新をおこなった。

- 議論時に自身が占い師であると宣言したエージェント数によるゲーム状態の制限
- 議論時における発言の種類の制限
- 占い師が人狼を発見した場合の行動の制限
- ゲーム開始直後 (0 日目) の夜における占い師役職の占い先の選択方法の制限
- 議論時にエージェントは同時発言するという制限

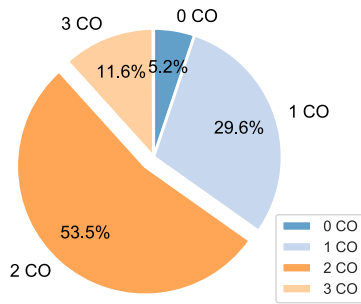


図 3 1 日目議論開始時の占い師 CO エージェント数の割合
Fig. 3 Percentage of agents who declared “SEER CO” at the start of the discussion on Day1

表 2 2CO のゲーム状態における実際の役職の内訳
Table 2 Actual roles at game states of “2 CO”

2CO における役職の内訳	回数
占い師と裏切り者	4,588
その他	755

3.6.1 議論時に自身が占い師であると宣言したエージェント数によるゲーム状態の制限

占い師役職は人狼ゲームにとって重要な役職であり、人狼陣営のプレイヤーが占い師役職を騙ることは一般的な作戦である。そのため、人狼ゲームでは占い師役職として振る舞うプレイヤーが複数存在するゲーム展開が多いと考えられる。したがって、本研究では自身が占い師役職であると宣言したプレイヤー数が一定のゲーム展開のみ考慮する。

人狼知能プロジェクトが提供している 2020 年におこなわれた人狼知能大会のログから、1 日目の議論開始時に占い師 CO をおこなったエージェントの数の割合を図 3 に示す。ここで、「CO (COMINGOUT)」とは自身の役職について宣言する意味の発言である。また、図 3 では、占い師 CO をおこなったエージェントが N 名いたことを「 N CO」と表している。

図 3 より、占い師 CO をおこなうエージェントが 2 体の場合が 53.5%と最も多く、人狼知能大会におけるゲーム展開の半数以上を占めている。そのため、占い師 CO をおこなうエージェントが 2 体のゲーム展開が人狼知能大会で最も多いと考えられる。

さらに、占い師 CO エージェントが 2 体のゲームにおける、占い師 CO をおこなったプレイヤーの実際の役職の内訳を表 2 に示す。表 2 より、占い師役職と裏切り者役職が占い師 CO をおこなっているゲーム展開が 8 割以上であることが分かる。この結果から、「2 CO」の中で、占い師役職と裏切り者役職が占い師 CO しているゲーム展開が最も多いと考えられる。

表 3 2 日目の議論開始時における占い師役職と裏切り者役職の発言の種類と回数

Table 3 Type and frequency of utterances for SEER and POSSESSED at the start of the discussion on Day2

発言種類	回数
占い結果報告発言	2,459
COMINGOUT 発言	1,573
その他	1,283

表 4 COMINGOUT 発言の内訳

Table 4 Detail of COMINGOUT utterances

COMINGOUT 発言の内容	回数
COMINGOUT POSSESSED (裏切り者)	660
COMINGOUT SEER (占い師)	472
COMINGOUT WEREWOLF (人狼)	441

簡易化した 5 人狼ゲームでは、人狼知能大会において最も多い状態だと考えられる占い師役職と裏切り者役職が占い師 CO をおこなうゲームのみ考慮する。

3.6.2 議論時における発言の種類の制限

人狼ゲームでは多様な発言が可能である。しかし、すべての発言を考慮すると状態数が非常に多くなると考えられる。そのため議論時における発言の種類を制限する。

1 日目では、人狼ゲームにおいて重要な情報である占い師 CO を発言したプレイヤーの占い結果報告発言のみを考慮する。占い結果報告発言とは占い師 CO を発言したプレイヤーが、占った結果そのプレイヤーは人間だ、あるいは人狼だということを報告するための発言である。

先行研究 [9] では、2 日目の議論において占い師 CO を発言したプレイヤーの占い師結果報告または、発言をしない行動を考慮していた。さらなる調査から、人狼知能による人狼ゲームに近くなるように更新をおこなった。

本研究における 2 日目では、人狼知能大会のログから集計した傾向をもとに、占い師 CO を発言したプレイヤーによる占い結果報告発言、またはパワープレイを意図した CO 発言を考慮した。ここでパワープレイとは、ゲームに生存している人狼陣営のプレイヤーが村人陣営のプレイヤーより多くなった場合に、人狼陣営のプレイヤーが協調して村人陣営を排除することで勝利を狙うという戦術である。また、村人陣営のプレイヤーが人狼陣営の役職を騙ることで、パワープレイを妨害する戦術もある。

表 3 より、2 日目の議論時における最初のターンで占い師役職と裏切り者役職のエージェントの発言の種類と回数について示す。表 3 では、占い結果報告が全体の半分以上しか発言されていないことが分かる。次点で CO 発言が多いという結果となった。

この CO 発言について表 4 より、自身が占い師役職、裏切り者役職、人狼役職であると主張するエージェントが多いことが分かる。これは、人狼知能大会において、2 日目

にパワープレイ戦術やその妨害をおこなうエージェントが多数いるためだと考えられる。そのため2日目では、占い師 CO プレイヤの占い結果報告発言の他に、パワープレイやその妨害を意図した CO 発言を考慮する。

3.6.3 占い師が人狼を発見した場合の行動の制限

占い師役職は、自身の能力により任意のプレイヤの種族を知ることができる。村人陣営が勝利するためには人狼役職のプレイヤを投票によりゲームから排除しなければならない。したがって、能力により人狼役職を発見した場合は人狼役職のプレイヤに投票すべきだと考えられる。そのため、簡易化した5人狼ゲームでは、占い師役職が人狼役職を発見した場合は、そのプレイヤに対して「人狼だった」という占い結果報告発言をおこない、そのプレイヤに投票する。

3.6.4 議論時にエージェントは同時発言するという制限

2020 年からの人狼知能大会で用いられている人狼知能プラットフォームでは、ターン内で各エージェントに発言の順番がランダムに与えられて、その順番通りに発言するという仕様となっている。これにより、相手の発言を考慮した上で自身の発言を決定できる。しかし、同ターン内で相手の発言を考慮すると、同時発言と比べて状態数が非常に多くなる。そのため、過去の大会でも用いられていた同時発言の仕様を用いる。

3.6.5 ゲーム開始直後(0日目)の夜における占い師役職の占い先の選択方法の制限

人狼知能プラットフォームでは、0日目に夜のフェーズのみおこなわれる。また、0日目の夜フェーズでは占い師役職のみ能力を行使できる。しかし、0日目は昼のフェーズがなく議論がおこなわれず、行動を決定するための情報が得られない。そのため、簡易化した5人狼ゲームでは、0日目の占い師役職の占い先選択についてはランダムに決定される。

4. 実験と評価

4.1 実験の目的と方法

多人数ゲームの結果は相手の戦略に依存するため、戦略を評価するには複数の相手と対戦することが必要である。MCCFR は自己対戦で学習するため、学習した戦略は対戦相手に依存しない。そのため、得られた戦略は対戦相手によらず一定の強さを持つことが期待される。

本実験では実験エージェントとして、MCCFR アルゴリズムにより学習した戦略にしたがう「MCCFR エージェント」と、学習前の初期戦略にしたがう「RANDOM エージェント」を作成する。そして、各実験エージェントごとに人狼知能大会の複数のエージェントと対戦し、MCCFR アルゴリズムの有効性を評価する。

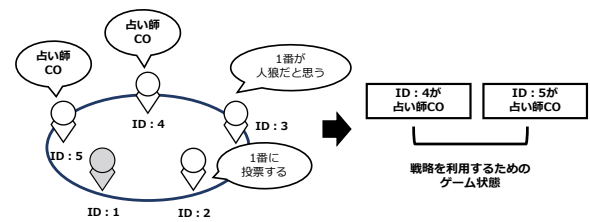


図 4 戦略を利用するためのゲーム状態の作成の例

Fig. 4 Example of creating a game state for using strategies

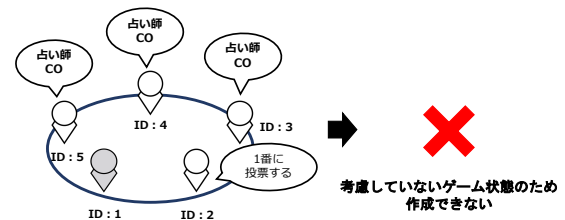


図 5 戦略を利用するためのゲーム状態が作成できない例

Fig. 5 Example of not being able to create the game state for using strategies

4.2 実験エージェント

4.2.1 MCCFR エージェント

MCCFR エージェントは、MCCFR アルゴリズムにより得られた戦略に従うエージェントである。このエージェントは、簡易化した5人狼ゲームで学習した戦略を実際の5人狼ゲームに対して利用する。

戦略は簡易化した5人狼ゲームの上で学習されており、実際の5人狼ゲームにおける状態を、簡易化した5人狼ゲームのものに合わせる必要がある。そのため MCCFR エージェントは、実際の5人狼ゲームをプレイしている際に「戦略を利用するためのゲーム状態」を作成する。

図4では、実際の5人狼ゲームをプレイしている際に戦略を利用するためのゲーム状態を作成する例について示す。簡易化した5人狼ゲームでは占い師 CO を発言するエージェントが2体である場合のゲーム展開についてのみ考慮している。そのため MCCFR エージェントは、占い師 CO を発言したエージェントから、戦略を利用するためのゲーム状態を作成する。このとき、他のプレイヤが学習時に考慮していない発言をおこなった場合、それらの発言は無視する。

実際のゲームにおけるゲーム展開が学習時に考慮していない展開となった場合、戦略を利用するためのゲーム状態が作成できない。図5では、戦略を利用するためのゲーム状態が作成できない例を示す。図5では、占い師 CO を発言するエージェントが3体いるため、学習時に考慮していないゲーム展開となっている。そのため、戦略を利用するためのゲーム状態は作成しない。

実際の5人狼ゲームにおいて戦略を利用するためのゲーム状態を作成できる場合は戦略にしたがい行動を決定

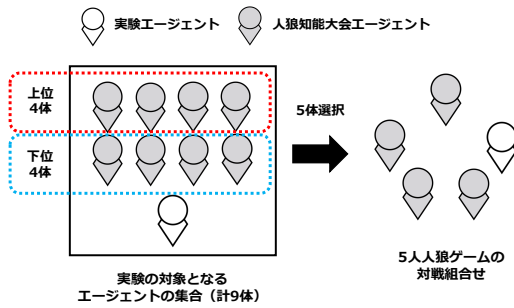


図 6 複数の人狼知能大会エージェントとの 5 人対戦

Fig. 6 5-player games with multiple AI Werewolf Competition agents

する。反対に、戦略が利用できない場合はランダムに行動を決定する。

また人狼知能大会では、大多数のエージェントが自身の投票対象を報告する発言をおこなう。そのため MCCFR エージェントも同様に、戦略により決定した投票対象について報告する発言をおこなうこととした。

4.2.2 RANDOM エージェント

RANDOM エージェントは MCCFR アルゴリズムの初期戦略として表現できるすべての行動が等確率で決定される戦略にしたがうエージェントである。このエージェントは、MCCFR アルゴリズムによって学習した戦略の効果を比較するために作成する。

4.3 実験条件

図 6 のように、各実験エージェントごとに過去大会の上位・下位各 4 体のエージェントと対戦をおこなう。これは、実験エージェントが人狼知能大会でどれだけ有効であるかについて確認するためである。合計 9 体のエージェントから、5 人対戦のすべての組合せで 100 ゲームを 1 セットとして 100 セット対戦をおこなう。

対戦に利用する過去の人狼知能大会エージェントは、2018 年に開催された第 4 回人狼知能大会（以降、2018 年大会）と、2019 年に開催された第 1 回国際人狼知能大会（以降、2019 年大会）を用いる。2018 年大会と 2019 年大会は、本研究における簡易化した 5 人狼ゲームにも用いている、エージェントが同時に発言するという仕様を用いていた大会である。また戦略は、MCCFR の反復回数 5 千万回時点のものを用いる。

4.4 実験結果と考察

各年度の上位・下位各 4 体の大会エージェントと実験エージェント合計 9 体から、5 人対戦のすべての組合せに対して 100 セットおこなった対戦による実験エージェントの合計勝率を表 5 に示す。また表 6 に、各大会年度の対戦環境において実験エージェントが MCCFR エージェントである場合に、大会エージェントの合計勝率における最大

表 5 実験エージェントが参加した 5 人対戦の組合せにおける勝率の合計

Table 5 Total win rate for each experimental agent

	RANDOM	MCCFR
2018 年大会の対戦環境	43.91%	48.51%
2019 年大会の対戦環境	44.76%	47.36%

表 6 実験エージェントが MCCFR エージェントである場合の大会エージェントの最大勝率と最小勝率

Table 6 Maximum and minimum win rate of AI Werewolf Competition agents in environments with MCCFR agent

	最大勝率	最小勝率
2018 年大会の対戦環境	61.98%	49.43%
2019 年大会の対戦環境	61.19%	47.13%

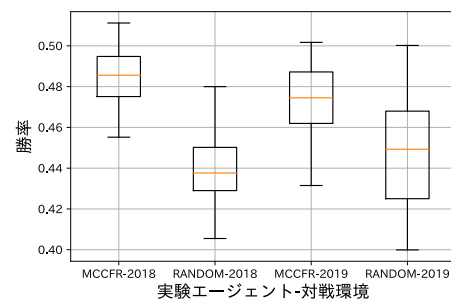


図 7 各実験エージェントの勝率の箱ひげ図

Fig. 7 Box plot of the win rate of each experimental agent

勝率と最小勝率について示す。

表 5 より、MCCFR エージェントの合計勝率が RANDOM エージェントと比べて向上していることが分かる。そして、表 5 の MCCFR エージェントの合計勝率と表 6 の最小勝率を比較すると、MCCFR エージェントの勝率が大会エージェントの最小勝率と比べて近い値または上回る値となった。人狼知能大会のエージェントは予選を通過した十分強いエージェントだと考えられるため、MCCFR エージェントの合計勝率は低くないといえる。しかし、表 5 の MCCFR エージェントの合計勝率が、表 6 の最大勝率と比べて低いことから現状の戦略では人狼知能大会で上位の順位を取ることは困難だと考えられる。

図 7 に、各実験エージェントが参加した 5 人対戦の各組合せに対して 100 セットおこなった対戦による合計勝率の箱ひげ図を示す。この図から複数のエージェントとの対戦において MCCFR エージェントは、RANDOM エージェントと比べて全体的に勝率が向上していることが分かる。したがって、MCCFR アルゴリズムにより人狼ゲームにおいて勝率を向上させる戦略を学習できたと考えられる。

また図 8 に、各年度の対戦環境において、実験エージェントが参加している 5 人対戦の各組合せに対して 100 セットおこなった対戦における MCCFR エージェントと RANDOM エージェントとの勝率の差をヒストグラムに示す。

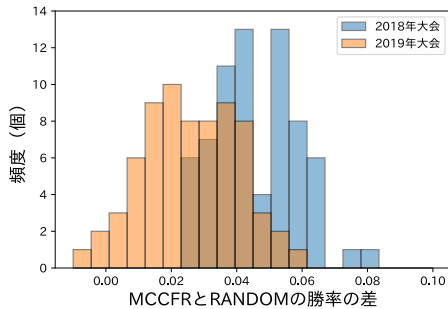


図 8 各対戦組合せにおける MCCFR エージェントと RANDOM エージェントの 100 セットの勝率の差

Fig. 8 Difference in win rate of 100 sets between MCCFR agent and RANDOM agent for each 5-player game

表 7 学習により勝率が減少した対戦組合せにおける実験エージェントの勝率

Table 7 Win rate of experimental agents in the 5-player game with decreased win rate

エージェント	勝率
RANDOM エージェント	47.53%
MCCFR エージェント	46.65%

図 8 より学習によって全体的に勝率が上昇していることが分かる。また、各大会年度の対戦環境における結果を比較すると、2018 年大会の対戦環境と比べて 2019 年大会の対戦環境では勝率の上がり幅が小さいことが分かる。これは、2019 年大会では 2018 年大会と比べて、個々のエージェントの戦略に多様性があるためであるためと考えられる。そのため、2019 年大会のエージェントとの対戦では、2018 年大会におけるエージェントとの対戦と比べて勝率を上昇させることが難しくなっていると考えられる。

図 8 より、2019 年大会の対戦環境では学習により全体的に勝率が上昇しているものの、いくつかの 5 人対戦の組合せでは勝率が変動していない、または僅かに減少しているものが見られた。この特徴的な対戦組合せについて、勝率が最も減少した対戦組合せにおける実験エージェントの 100 セットの勝率を表 7 に示す。表 7 より、MCCFR エージェントの勝率が RANDOM エージェントと比べて 0.88 ポイント減少した。この結果から、対戦相手によっては有効ではない場合があることを確認した。

本研究による簡易化した 5 人狼ゲームにおいて学習した戦略では、実際のゲーム状態での適応能力が十分でなく、相手の戦略によっては有効ではない場合が見られた。したがって、汎用的な戦略を得るために、簡易化した 5 人狼ゲームの改善が必要である。

5. まとめ

本研究では、5 人狼ゲームに対して MCCFR アルゴリズムの適用をおこなった。MCCFR アルゴリズムを適用す

るために簡易化した 5 人狼ゲームを構築し、その上で戦略を学習した。また、人狼ゲームにおける MCCFR アルゴリズムの有効性の検証のために、複数の人狼知能大会エージェントと対戦した。

対戦実験の結果から、MCCFR アルゴリズムで学習した戦略にしたがうことで、複数の人狼知能大会エージェントとの対戦組合せにおいて全体的に勝率を向上させることができた。したがって、MCCFR アルゴリズムが人狼ゲームに対して有効であることを示した。

しかし、本研究におけるゲームモデルで学習した戦略では、実際のゲーム状態での適応能力が十分でなく、一部のエージェントの戦略によっては有効でないことが確認された。そのため、より汎用的な戦略を得るために簡易化した 5 人狼ゲームを改善することが必要だと考えられる。

参考文献

- [1] 鳥海不二夫, 片上大輔, 大澤博隆, 稲葉通将, 篠田孝祐, 狩野芳伸: 人狼知能: だます・見破る・説得する人工知能, 森北出版 (2016).
- [2] Moravčík, M., Schmid, M., Burch, N., Lisý, V., Morrill, D., Bard, N., Davis, T., Waugh, K., Johanson, M. and Bowling, M.: DeepStack: Expert-level artificial intelligence in heads-up no-limit poker, *Science*, Vol. 356, No. 6337, p. 508–513 (online), DOI: 10.1126/science.aam6960 (2017).
- [3] Brown, N. and Sandholm, T.: Superhuman AI for heads-up no-limit poker: Libratus beats top professionals, *Science*, Vol. 359, No. 6374, pp. 418–424 (online), DOI: 10.1126/science.aao1733 (2018).
- [4] Zinkevich, M. et al.: Regret minimization in games with incomplete information, *Advances in neural information processing systems*, pp. 1729–1736 (2008).
- [5] Abou Risk, N. and Szafron, D.: Using counterfactual regret minimization to create competitive multiplayer poker agents., *AAMAS*, pp. 159–166 (2010).
- [6] Brown, N. and Sandholm, T.: Superhuman AI for multiplayer poker, *Science*, Vol. 365, No. 6456, pp. 885–890 (online), DOI: 10.1126/science.aay2400 (2019).
- [7] Lanctot, M., Waugh, K., Zinkevich, M. and Bowling, M. H.: Monte Carlo Sampling for Regret Minimization in Extensive Games., *NIPS*, pp. 1078–1086 (2009).
- [8] 畢 曉恒, 田中哲朗: 対話のない人狼ゲームの戦略, ゲームプログラミングワークショップ 2015 論文集, Vol. 2015, pp. 25–30 (2015).
- [9] 小島大征, 内種岳詞, 岩田員典, 伊藤暢浩: 5 人狼ゲームにおける MCCFR アルゴリズムの適用, 日本知能情報ファジィ学会 ファジィ システム シンポジウム講演論文集 第 37 回ファジィシステムシンポジウム, 日本知能情報ファジィ学会, pp. 507–512 (2021).
- [10] Neller, T. W. and Lanctot, M.: An introduction to counterfactual regret minimization, *Proceedings of Model AI Assignments, The Fourth Symposium on Educational Advances in Artificial Intelligence (EAAI-2013)* (2013).