人狼ゲームにおける最適な発言のゲーム理論による分析

清水 大輔　長谷部 浩二

人狼ゲームは，プレイヤーにとって一部の情報が隠されている不完全情報ゲームの一種である．本研究では，ゲーム理論に基づき，プレイヤー 3 人および 5 人の人狼ゲームにおいて各役職のプレイヤーがすべき最適な発言について分析する．そのために，まず人狼ゲームにおいて各プレイヤーの役職が隠されている状況をベイジアンゲームとして定式化する．その上で，相手の発言を聞いたときのプレイヤーの投票先の決め方を定義し，自らの信念に基づいて計算される期待利得を最大化するような戦略として最適な発言を求める．分析の結果，プレイヤー 3 人のゲームにおいては，支配戦略または弱支配戦略となる発言，すなわち相手のあらゆる発言に対して最適となる発言が役職ごとに存在することが明らかとなった．また，5 人のゲームにおいては，狂人に関して支配戦略となる発言が存在したほか，狂人以外の役職に関しては，相手の発言のパターンによって最適な発言が変わることが観察された．

1　研究の背景と目的

近年，ゲームをプレイするプログラムの研究開発が活発に行われている．チャスや囲碁・将棋などのように，各プレイヤーが盤面の情報を全て知ることができる完全情報ゲームについては，2016 年に Google DeepMind 社の囲碁プログラム「アルファ碁」が韓国のプロ棋士に勝ち越すなど，人間の能力を大きく超える強さのプログラムが開発されている．これに対して，麻雀やポーカーなどのように，プレイヤーにとって一部の情報が隠されているゲームは不完全情報ゲームと呼ばれる．不完全情報ゲームに関してもさまざまな研究が行われているが，まだ人間に勝利するレベルには達していないものが多い．

こうした不完全情報ゲームをプレイするプログラムに関する研究分野において近年注目されているゲームの一つに，人狼ゲームがある [1]．これは，数人から十数人のプレイヤーで行う多人数ゲームであり，プレイヤー同士の会話などによって進行するコミュニケーションゲームとしての側面も持つ．村人陣営と人狼陣営の 2 つの陣営（チーム）に分かれ，互いに相手陣営のプレイヤーをゲームから排除していくことで勝利を目指すゲームである．各プレイヤーには自分以外のプレイヤーの所属する陣営が知らされないため，不完全情報ゲームに分類される．

人狼ゲームをプレイするプログラムの研究開発を目的とした「人狼知能プロジェクト」が 2013 年に発足した [1]．また，2015 年からプログラムどうしを対戦させる大会が毎年開かれており [2]，2019 年には初めての国際大会が開催されている [3]．

こうした中で，ゲームの戦略に関する手法も数多く提案されている．大澤ら [4] は，人狼，占い師，村人の 3 人でプレイされる人狼ゲームにおいてどのような戦略（発言）があり得るのかを検討している．3 人が同時に 1 回のみ発言したあと投票を行うという単純化された設定のゲームにおいて，ゲームの状態を可能世界の組み合わせで表し，各プレイヤーが行動の選択肢をどのくらい削減できるかを分析してお

---

* Game Theoretic Analysis of Best Utterances in Werewolf Game

Daisuke Shimizu, 筑波大学 情報理工学位プログラム, Master's Program in Computer Science, University of Tsukuba.

Koji Hasebe, 筑波大学 システム情報系, Faculty of Engineering, Information and Systems, University of Tsukuba.

り，その結果，ある条件のもとで占い師や人狼のプレイヤーが発言の選択肢を削減できることが示された．梶原ら [5] は，サポートベクターマシン（SVM）を用いてプレイヤーの役職を推定する方法を提案している．梶原らは，まず，村人陣営のプレイヤーが人狼が誰であるかを正しく予測することが村人陣営の勝率を向上させることを示した．次に，SVM を用いて人狼を推定する手法を提案し，2015 年に開催された第 1 回人狼知能大会の決勝に出場したチームのエージェントより高い性能で人狼を推定できることを示した．また，大川ら [6] は，梶原らが採用した特徴にさらに 3 つの特徴を追加し，深層学習を用いて役職を推定する手法を提案している．大川らは，追加した特徴の有効性を正答率の向上によって示し，ランダムに行動を選択するエージェントより学習モデルを用いたエージェントのほうが勝率が高くなることを報告した．このようにさまざまな研究がなされているが，一方で，人狼ゲームにおいて自分の所属する陣営の勝利のためにふさわしい発言を選ぶための体系的な方法は未だ確立されていない．

そこで，本研究は，人狼ゲームにおいてプレイヤーが自分の所属する陣営の勝利のためにどのような発言をすべきかを分析することを目的とする．そのためのアプローチとして，ゲーム理論における不完備情報ゲームの一種であるベイジアンゲーム [7] の概念を用いる．ベイジアンゲームとは，他のプレイヤーの利得などに関する情報がプレイヤー間で共有されていないゲームであり，人狼ゲームにおいて自分以外のプレイヤーの役職が隠されている状況は，自分以外のプレイヤーの利得関数が不確実であるベイジアンゲームとして記述することができる（ベイジアンゲームの詳細については文献 [8] などを参照のこと）．その上で，ゲームの均衡を構成する最適反応を求めるために，次のことを行う．プレイヤーの投票行動について，妥当と考えられる投票先の決定方法を，全てのプレイヤーにとっての共有知識としてあらかじめ定義する．さらに，自分が選択できる発言を，自らの信念に基づいて計算される期待利得によって比較することで，所属する陣営の勝利のために各プレイヤーがどのような発言をすればよいかを分析する．

以上の分析の結果として，プレイヤー 3 人のゲームにおいては，支配戦略または弱支配戦略となる発言，すなわち相手のあらゆる発言に対して最適となる発言が人狼，占い師，村人の役職ごとに存在することが明らかとなった．また，5 人のゲームにおいては，狂人に関して支配戦略となる発言が存在したほか，狂人以外の役職に関しては，相手の発言のパターンによって最適な発言が変わることが観察された．実際のゲームのプレイにおいては，自分の発言を選択する直前までのプレイヤーの発言を聞いた上で，その内容に応じて最適な発言を選択することができると考えられる．

本論文の構成は以下のとおりである．第 2 章では，人狼ゲームの一般的なルールを説明する．第章では，最適な発言を求めるためのモデルの概要を述べる．第 4 章および第 5 章では，プレイヤー 3 人および 5 人のゲームにおける最適な発言の分析の結果について述べる．最後に第 6 章では，結論と今後の課題について述べる．

2　人狼ゲームのルール

この章では，人狼ゲームの一般的なルールを説明する．人狼ゲームは，自然言語によるコミュニケーションによって進行する対話型ゲームであり，数人から十数人のプレイヤーで行われる．プレイヤーは村人陣営と人狼陣営の 2 つの陣営に分かれ，それぞれの陣営の勝利条件を満たすことを目的とする．人狼ゲームのルールにはさまざまなバリエーションが存在するが，本研究では，人狼知能プロジェクトが開催する人狼知能大会において採用されている標準的なルール [9] に従うものとする．

2.1　ゲームの進行

ゲーム開始時に，各プレイヤーに 1 人 1 つの役職が割り振られる．このとき割り振られた役職により，プレイヤーは村人陣営と人狼陣営に分かれる．村人と占い師は村人陣営に所属し，人狼と狂人は人狼陣営に所属する．ここで，自分の役職は他のプレイヤーには公開されない．すなわち，各プレイヤーは自分の役職しか知ることができない．

ゲームは，昼と夜の 2 つのフェーズから構成され

る．0 日目の夜から始まり，それ以降，昼と夜が繰り返される．昼のフェーズでは，生存しているプレイヤー全員が参加して対話が行われたのち，投票による多数決で 1 人のプレイヤーが追放される．夜のフェーズでは，人狼により 1 人のプレイヤーが襲撃される．ゲームが進むにつれ，投票による追放や人狼の襲撃によってプレイヤーがゲームから排除されていく．いずれかの陣営が勝利条件を満たしたら勝者が決まり，ゲームが終了する．村人陣営の勝利条件は，人狼を全滅させることである．狂人は，人狼陣営に所属するが人狼ではないため，残存していてもよい．人狼陣営の勝利条件は，人間の数を人狼の数以下にすることである．このとき,「人間」とは，村人陣営のすべてのプレイヤーと狂人を表す．人狼陣営に属する狂人も「人間」として数えることに注意する．仮に自らがゲームから排除されていたとしても，所属する陣営が勝利していれば，自らも勝者に含まれる．

2.2 プレイヤーの役職

プレイヤーは，割り当てられた役職により，ゲームの中でさまざまな能力や目的が与えられる．以下でその特徴の概要を述べる．

2.2.1 占い師

占い師は，プレイヤーを占うことができる能力を持つ．すなわち，夜のフェーズで 1 日に 1 人を指定して，そのプレイヤーが人狼であるか否かを知ることができる．誰を占うかは，昼のフェーズの議論なども考慮して占い師本人が決定する．もし，占いによって，あるプレイヤーが人狼であると分かれば，そのプレイヤーを追放することで村人陣営の勝利に貢献できる．ただし，占いの結果を知らされるのは，占い師本人に対してのみである．

2.2.2 村人

村人は，何の特殊能力も持たない．昼のフェーズの対話の内容や，各プレイヤーの投票行動，誰が襲撃されたかなど，全体に公開されている情報をもとにして，人狼が誰であるかを推測し，人狼を追放するために行動する．

2.2.3 人狼

人狼は，プレイヤーを襲撃することができる能力を持つ．すなわち，夜のフェーズで 1 日に 1 人のプレイヤーを選んで，そのプレイヤーをゲームから排除できる．人狼が自らを襲撃することはできないものとする．人狼がすべて追放されると人狼陣営の敗北となるため，対話において自らの正体を隠してうまく振る舞う必要がある．

2.2.4 狂人

狂人は，何の特殊能力も持たない．人狼陣営に属するが人狼ではない．勝利条件の判定に際しては「人間」として数えられる．人狼陣営の勝利条件を満たすことを目指すので，人狼が村人陣営に正体を隠したり，村人陣営を騙したりするための手助けをする．ただし，狂人自身は誰が人狼であるかを知らないため，人狼が誰であるかを推測して行動する必要がある．

3 最適な発言を求めるためのモデル

本研究で行う分析の概要は以下の通りである．人狼ゲームにおいては，投票により追放されることになったプレイヤーの役職と自身の役職に応じて各プレイヤーの利得が決定するため，各プレイヤーの役職が隠されている状況は，各プレイヤーの利得関数が不確実であるベイジアンゲームとして記述できる．そこで，まず，投票結果が各プレイヤーにもたらす利得をベイジアンゲームとして定式化する．その上で，各プレイヤーが自分以外のプレイヤーの発言を聞いたときの投票先の決定方法を定義する．この決定方法は全てのプレイヤーが従うものと仮定し，また全てのプレイヤーにとって共有知識であると仮定する．これにより，各プレイヤーが相手の発言を聞いた上で，最も高い利得をもたらす発言を戦略として選択する．

3.1 ベイジアンゲームゲームによる定式化

各プレイヤーが自身の戦略として発言を選ぶという点に着目するとき，プレイヤー n 人の人狼ゲームは，次のような要素をもつベイジアンゲーム $G = (N, A, \Theta, p, u)$ として定式化できる．ただしここで，G の各構成要素は以下のように定義される．

- プレイヤーの集合 $N = \{1, \cdots, n\}$

- 各プレイヤー i の行動（発言）の集合
  $A_i = \{t_0, t_{-i,W}, t_{-i,V}\}$，$A = A_1 \times \cdots \times A_N$．
  ただし，t_0 は「私は占い師ではない」，$t_{j,W}$ は「私は占い師だ．占いの結果，プレイヤー j は人狼だった」，$t_{j,V}$ は「私は占い師だ．占いの結果，プレイヤー j は人狼ではなかった」をそれぞれ意味する．
  $t_{-i,W} = \{t_{1,W}, \cdots, t_{i-1,W}, t_{i+1,W}, \cdots, t_{N,W}\}$，
  $t_{-i,V} = \{t_{1,V}, \cdots, t_{i-1,V}, t_{i+1,V}, \cdots, t_{N,V}\}$．
- 各プレイヤー i のタイプ（役職）の集合
  $\Theta_i = \{$人狼，占い師，狂人，村人$\}$（プレイヤー 5 人のとき）．$\Theta = \Theta_1 \times \cdots \times \Theta_N$．
- タイプの事前分布 $p\colon \Theta \to [0,1]$
  ただし，任意の $\theta \in \Theta$ について，$p(\theta) = \dfrac{1}{|\Theta|}$．
- 利得関数 $u\colon A \times \Theta \to \{\pm 1, 0\}$，$u = (u_1, \cdots, u_N)$
  ただし，任意の戦略プロファイル $a \in A$ とタイププロファイル $\theta \in \Theta$ について，プレイヤー i が勝利陣営のプレイヤーであれば $u_i(a, \theta) = 1$，敗北陣営のプレイヤーであれば $u_i(a, \theta) = -1$，勝負がついていなければ $u_i(a, \theta) = 0$．

上記の定式化において，プレイヤーのタイプが役職を表す．また，プレイヤーの行動とは，各プレイヤーが選択できる発言内容を表す．ここでは分析を単純化するため，各プレイヤーが発言できる内容は t_0 と $t_{-i,W}$ に限る．すなわち，各プレイヤーは，占い師として名乗り出ないか，占い師として名乗り出て 1 人のプレイヤーに関して「占いの結果，人狼だった」と発言するかのいずれかを選択するものとする．

3.2 投票先の決定方法

まず，各プレイヤーが自分以外のプレイヤーの発言を聞いたときの投票先の決め方を定義する．各プレイヤーは，自分以外のプレイヤーの役職に関して，可能世界に基づいて考えるものとする．ここで可能世界とは，n 人のタイプ（役職）の組み合わせのことであり，1 つの可能世界は 1 つのタイププロファイル $\theta \in \Theta$ によって表される．各プレイヤーは信念として，各可能世界がプレイされている確率（確からしさ）を考えており，プレイヤー i の信念を確率分布 $p_i\colon \Theta \to [0,1]$ で書くことにする．例えば，プレイヤー 2 の役職が占い師であるとき，プレイヤー 2 の役職が人狼であるような可能世界 θ' にプレイヤー 2 が割り当てる確率は 0 であるから，$p_2(\theta') = 0$ となる．

各プレイヤーは，相手の発言を聞くとき，以下の項目に従って可能世界を絞り込んでいく．

- もし自分が占い師ならば，既に得ている占いの結果に矛盾する可能世界を削除する．
- 自分以外のプレイヤーの発言について，
  - 自分の役職や占い結果など，自分の持っている確かな情報に矛盾する発言は無視する．
  - 無視しない発言は，事実として素直に受け取る．
  - 無視しない発言同士が矛盾する場合，それぞれの可能性を等しく考える．

各プレイヤーは，以上の項目に従って可能世界を絞り込んだのち，残っている可能世界に基づいて自分以外のプレイヤーが人狼である可能性，占い師である可能性，狂人である可能性，村人である可能性をそれぞれ求める．たとえば，あるプレイヤーが人狼である可能性は，残っている可能世界のうち当該プレイヤーが人狼である世界の確からしさの和として求める．

村人陣営のプレイヤーであれば人狼や狂人に投票することが望ましいため，投票先は，各プレイヤーの人狼または狂人である可能性の割合に比例して確率を割り振る混合戦略として定める．人狼陣営のプレイヤーであれば占い師に投票することが望ましいため，投票先は，各プレイヤーの占い師である可能性の割合に比例して確率を割り振る混合戦略として定める．

3.3 最適な発言を求めるアルゴリズム

各プレイヤー i は，自分以外のプレイヤーの発言 t_{-i} が与えられたとき，最適な発言として，自らの信念 p_i をもとに計算される期待利得を最大にするような発言を選ぶ．ここで，プレイヤー i が戦略（発言）t_i を選ぶときの期待利得 $EU_i(t_i, t_{-i})$ とは，次の式で定められるものとする．

$$EU_i(t_i, t_{-i}) = \sum_{k=1}^{m} p_i(\theta^{(j)}) u_i(t_i, t_{-i}; \theta^{(j)})$$

上式における m は可能世界の個数であり，$m = \dfrac{1}{|\Theta|}$ である．また，j 番目の可能世界をタイププロファイ

ル $\theta^{(j)}$ として表してある．なお，削除された可能世界の確からしさは 0 と考えることにする．

4 プレイヤー 3 人のゲームの分析

4.1 ゲームのルール

ここでは，大澤ら [4] によって提案されたプレイヤー 3 人の人狼ゲームを対象とする．これは人狼ゲームの最少人数の形式であり，人狼，占い師，村人のうちいずれか 1 つの役職を持つ 3 人のプレイヤーでプレイされる．占い師と村人は村人陣営，人狼は人狼陣営に属する．各プレイヤーは，自分の属する陣営の勝利を目指して行動する．ゲームの流れを以下に示す．

1. はじめに，人狼，占い師，村人の役職が 3 人に割り振られる．このとき，役職は当人にのみ知らされ，自分以外の役職は知ることができない．
2. 占い師となったプレイヤーは，対話のフェーズが始まる前に自分以外のプレイヤーを 1 人選び，そのプレイヤーが人狼であるか否かを知ることができる．プレイヤー 3 人のゲームでは占い師の他にプレイヤーが 2 人しかいないため，この占いによって誰が人狼であるのかを確実に知ることができる．
3. 対話のフェーズでは，各プレイヤーが同時に 1 回のみ発言を行う．ここでは，各プレイヤー i が選択できる発言の種類を次の 3 種類に限っている．
   - t_0：「私は村人である」
   - $t_{j,W}$：「私は占い師である．占いの結果，プレイヤー j が人狼であった」(ただし，$j \neq i$)
4. 各プレイヤーは，全員の発言を聞いた上で，誰を追放したいかを投票によって表明する．多数決により，最も多く得票したプレイヤーが追放される．追放されたプレイヤーが人狼であれば村人陣営の勝利，占い師または村人であれば人狼陣営の勝利となる．最多得票者が複数人いた場合は，誰も追放されず，引き分けとなる．

4.2 各役職に関する最適な発言

村人，占い師，人狼のそれぞれの役職について，第 3 章に示したアルゴリズムにより求められる最適な発言を示し，結果に関しての考察を述べる．

4.2.1 村人

村人に関しては，自分が占い師であると詐称し，任意のプレイヤーについて「占いの結果，人狼だった」と発言することが弱支配戦略となった．

人狼は占い師に投票しようとするため，発言を聞いて自分以外のどちらが占い師であるかを判断することになる．人狼は，事実に矛盾しない占い結果を報告した自称占い師を占い師として見なすことになるため，人狼に本物の占い師が誰であるかを分からせないようにするためには，村人が自分のことを占い師であると詐称し，人狼に対して「占いの結果，人狼だった」と発言することが望ましい．ただし，村人が発言を選択する時点では，自分以外の 2 人のプレイヤーについて，どちらが人狼であるのか分からないため，占いの対象として各プレイヤーを確率 1/2 ずつで選択することになる．

4.2.2 占い師

事実を述べること，すなわち，占い師として名乗り出たうえで占いの結果を偽りなく伝えることが弱支配戦略となった．占い結果を偽りなく伝えることで，自分が占い師であることが人狼に知られてしまう恐れがあるが，人狼の正体を村人に伝えることも重要であるということだと考えられる．

4.2.3 人狼

人狼に関しては，自分が占い師であると詐称し，任意のプレイヤーについて「占いの結果，人狼だった」と発言することが支配戦略となった．人狼が勝利するためには，人狼と村人がともに占い師に投票することが必要となる．村人が占い師に投票するようにさせるためには，村人に対して「占いの結果，人狼だった」と言わないこと，すなわち，占い師に対して「占いの結果，人狼だった」と発言することが重要である．ただし，村人と同様に，発言の時点では，自分以外の 2 人のプレイヤーについて，どちらが占い師であるのか分からないため，占いの対象として各プレイヤーを確率 1/2 ずつで選択することになる．

5　プレイヤー 5 人のゲームの分析

第 3 章に示したアルゴリズムをプレイヤー 5 人の人狼ゲームに適用する．以下，第 5.1 節において 5 人のゲームにおける変更点を示し，第 5.2 節において結果を示す．

5.1　ゲームのルール

ここでは，人狼知能大会において採用されている 5 人のゲームのルール [9] に従う．このルールにおいてプレイヤー 3 人のゲームと異なる点を記す．5 人の役職は，村人 2 人，占い師 1 人，人狼 1 人，狂人 1 人となる．このうち占い師は，対話のフェーズの前に 1 人のプレイヤーを選んで，そのプレイヤーが人狼であるか否かを知ることができるが，プレイヤーが 5 人の場合は，いずれか 1 人のプレイヤーが「人狼でない」と分かったからと言って直ちに誰が人狼であるかを知ることはできない．本研究では，単純化のため，占われたプレイヤーが人狼だと分かったケースに限定する．これに伴い，各プレイヤー i が選択できる発言は，

- t_0：特に役職を名乗り出ない
- $t_{j,W}$：「私は占い師だ．占いの結果，プレイヤー i は人狼だった」（ただし，$j \neq i$）

の計 5 種類となる．

対話のフェーズ終了後の投票の結果，いずれか 1 人のプレイヤーが追放される．ただし，プレイヤー 3 人のゲームと同様に，最多得票者が複数人いる場合は，誰も追放されず，引き分けとなる．追放されたプレイヤーが存在するとき，そのプレイヤーが人狼であれば村人陣営の勝利となり，人狼でないときはその時点では勝負が決定しないため，引き分けとして全員の利得を 0 点と見なすこととする．

5.2　各役職に関する最適な発言

村人，占い師，人狼，狂人のそれぞれの役職について，上記アルゴリズムにより求められる最適な発言を示し，結果に関しての考察を述べる．

自分以外の 4 人のプレイヤーの発言の組み合わせは，$9^4 = 625$ 通りある．そのそれぞれに対して，自分のどの発言が最適であるかを調べた．分析の結果を，自分の役職が村人であるとき，占い師であるとき，人狼であるとき，狂人であるときの各場合に分けて述べる．

5.2.1　村人

村人の最適な発言として，次の 2 つのパターンが観察された．

- 自分以外の 4 人のなかに自称占い師がちょうど 2 人いて，かつ次の 3 つの条件をいずれも満たさないとき（625 通りのうち，12 通りある），かつそのときに限り，村人は，占い師として名乗り出て，自称占い師のどちらかに対して「占いの結果，人狼だった」と発言することが最適となる．
  - 条件 1．少なくとも 1 人の自称占い師が，自分（村人）に対して「占いの結果，人狼だった」と言っている．
  - 条件 2．2 人の自称占い師が，同一のプレイヤーに対して「占いの結果，人狼だった」と言っている．
  - 条件 3．少なくとも 1 人の自称占い師が，他方の自称占い師に対して「占いの結果，人狼だった」と言っている．
- それ以外のときは，t_0，すなわち特に何も言わないことが最適となる．

村人が役職を詐称して占い師として名乗り出るとき，人狼，狂人，占い師，村人のそれぞれに対して確率 1/4 で「占いの結果，人狼だった」と発言することになる．狂人に対して発言したとき，人狼・狂人・占い師からは発言を無視されるが，もう 1 人の村人は発言を聞いて狂人のことを占い師だと思う度合いが増す．一方で，人狼に対して発言したとき，人狼や狂人が自分（村人）のことを占い師だと思う度合いが増し，したがって自分に投票される確率が増えることになる．このように，村人が占い師として名乗り出ることによる利益より不利益のほうが大きいため，村人は基本的に何も言わないことが最適となるものと考えられる．

5.2.2　占い師

占い師は，発言を選択する前に占い結果として人狼が誰であるかを知らされる．この仮定のもとでの占い

師の最適な発言として，次の 2 つのパターンが観察された．

- 自分以外の 4 人の発言の組み合わせのうち，いくつかのケースについては，t_0，すなわち何も言わないことが最適となる（625 通りのうち，全部で 75 通りある）．たとえば次のような場合である．占い師である自分を S として，それ以外に自称占い師 A，B がいるとする．また，S は「A が人狼である」との占い結果を得ているとする．A が「私は占い師だ．占いの結果，S は人狼だった」，B が「私は占い師だ．占いの結果，A は人狼だった」と言っているとき，S は何も言わないことが最適となる．
- それ以外のときは，事前に得ている占い結果を偽りなく伝えることが最適となる．

上記のように，占い師が占い結果を偽りなく報告することが最適とならないケースも一定数あるものの，大部分のケースにおいては 3 人のゲームと同様に占い結果を偽りなく報告することが最適となることが明らかとなった．

5.2.3 人狼

人狼の最適な発言として，次の 2 つのパターンが観察された．

- 自分以外の 4 人がともに t_0，すなわち何も言わないことを選ぶとき（625 通りのうち，1 通りだけある），かつそのときに限り，占い師として名乗り出て，自分以外の任意のプレイヤーに対して「占いの結果，人狼だった」と発言することが最適となる．
- それ以外のときは，t_0，すなわち特に何も言わないことが最適となる．

人狼が役職を詐称して占い師として名乗り出るとき，確率 1/2 で村人，1/4 で占い師，1/4 で狂人に対して「占いの結果，人狼だった」と発言することになる．発言の対象が狂人であった場合，その発言を聞いた村人 2 人が狂人のことを人狼だと思う度合いが増す．発言の対象が占い師であった場合，その発言を聞いた村人 2 人については占い師のことを人狼だと思う度合いが増すが，狂人については人狼のことを占い師だと思う度合いが増してしまう．発言の対象が村人であった場合，発言の対象となった村人からは無視されるが，発言を聞いたもう 1 人の村人については当該村人のことを人狼だと思う度合いが増し，他方で発言を聞いた狂人については人狼のことを占い師だと思う度合いが増す．このように，人狼が占い師として名乗り出ることによって人狼陣営の利益になる点もあるものの，人狼陣営が被る不利益も存在するため，人狼は基本的に何も言わないことが最適となるものと考えられる．

5.2.4 狂人

狂人の最適な発言としては，625 通りすべての場合において，t_0，すなわち特に何も言わないことが最適となった．したがって，占い師として名乗り出ないことが狂人の最適な戦略である．

狂人が役職を詐称して占い師として名乗り出るとき，確率 1/2 で村人，1/4 で占い師，1/4 で人狼に対して「占いの結果，人狼だった」と発言することになる．村人または占い師に対して発言すると人狼陣営に有利となり，人狼に対して発言すると人狼陣営に不利になる．村人 A に対して発言すると，他方の村人 B がその発言を聞いて村人 A のことを人狼だと思う度合いが増す．占い師に対して発言すると，村人がその発言を聞いて占い師のことを人狼だと思う度合いが増す．人狼に対して発言してしまうと，村人がその発言を聞いて人狼のことを人狼だと思う度合いが増し，人狼がその発言を聞いて狂人のことを占い師だと思う度合いが増す．このように，人狼に対して発言したときに不利になる度合いが村人または占い師に対して発言したときに有利になる度合いよりも大きいため，何も言わないほうが期待利得が大きく，最適となる．

6 結論と今後の課題

本研究では，人狼ゲームにおける最適な発言について，ゲーム理論を用いて分析した．具体的には，人狼ゲームを不完備情報ゲームの一種であるベイジアンゲームとして定式化し，投票先の決定方法をモデル化した上で，期待利得に基づいてプレイヤーの最適な発言を求めた．分析の結果として，プレイヤー 3 人のゲームにおいては，支配戦略または弱支配戦略となる発言，すなわち相手のあらゆる発言に対して最適

となる発言が人狼，占い師，村人の役職ごとに存在することが明らかとなった．また，5 人のゲームにおいては，狂人に関して支配戦略となる発言が存在したほか，狂人以外の役職に関しては，相手の発言のパターンによって最適な発言が変わることが観察された．

今後の課題として，実際にこの手法を用いてゲームをプレイするエージェントを作成し，過去の人狼知能大会の決勝に出場したプログラムと対戦させて性能を評価することが挙げられる．そのために，本研究で分析したゲーム開始直後の発言のみならず，ゲームが終了するまでの複数のターンにわたる発言や人狼の襲撃先の選択などについて，最適な戦略の分析に取り組む予定である．また，少人数でのゲームだけでなく，プレイヤー 15 人のゲームに適用できるようにすることも重要な課題である．15 人のゲームでは発言の組み合わせや可能世界の数が非常に大きくなるため，状態を抽象化するなどして計算量を削減する必要があると考えている．

参 考 文 献

[1] 片上大輔，鳥海不二夫，大澤博隆，稲葉通将，篠田孝祐，松原仁（2015）「人狼知能プロジェクト（〈特集〉エンターテイメントにおける AI）」『人工知能』第 30 巻，第 1 号，pp. 65–73.

[2] Toriumi, F., Osawa, H., Inaba, M., Katagami, D., Shinoda, K., and Matsubara, H. (2017). AI Wolf Contest — Development of Game AI Using Collective Intelligence —. *Computer Games*, pp. 101–115.

[3] 稲葉通将，片上大輔，狩野芳伸，大槻恭士（2019）「人狼知能と不完全情報ゲーム」『人工知能』第 34 巻，第 6 号，pp. 876–880.

[4] 大澤博隆，佐藤健（2016）「3 者間人狼における戦略の検討」『2016 年度人工知能学会全国大会（第 30 回）論文集』

[5] 梶原健吾，鳥海不二夫，稲葉通将，大澤博隆，片上大輔，篠田孝祐，松原仁，狩野芳伸（2016）「人狼知能大会における統計分析と SVM を用いた人狼推定を行うエージェントの設計」『2016 年度人工知能学会全国大会（第 30 回）論文集』

[6] 大川貴聖，吉仲亮，篠原歩（2017）「深層学習を用いて役職推定を行う人狼知能エージェントの開発」『ゲームプログラミングワークショップ 2017 論文集』pp. 50–55.

[7] J. C. Harsanyi. Games with Incomplete Information Played by Bayesian Players, I–III. *Management Science* vol.14(3): pp. 159–183 (Part I), vol.14(5): pp. 320–334 (Part II), vol.14(7): pp. 486–502 (Part III), 1967/1968.

[8] Osborne, M. J., and Rubinstein, A. (1994). *A Course in Game Theory*, The MIT Press.

[9] 人狼知能プロジェクト（2020）“2nd International AIWolf Competition Regulation” 〈`http://aiwolf.org/en/2nd-international-aiwolf-contest`〉（2020 年 8 月 17 日閲覧確認）