

人狼ゲームにおける人間らしいエージェントの要素の 分析：エージェントの存在通知の有無による影響

An analysis of influences of information that an agent plays the werewolf games

高田和磨^{1*} 杉原太郎¹ 五福明夫¹
Kazuma Takata¹, Taro Sugihara¹, and Akio Gofuku¹

¹ 岡山大学大学院自然科学研究科
¹ Graduate School of Natural Science and Technology, Okayama University

Abstract: This study analyzes human factors to implement a human-like agent in werewolf games. A comparative experiment was conducted to reveal influences regarding the degree of disclosure of the agent existence; seven players and agents participated in four games. Communication logs were collected from the game logs. Impressions for each player was reported in questionnaires with five-point scales and detection of the agent were reported in free writing. Although the most players, who did not know that an agent played the games, did not be aware of the true character of the agent, the most players who knew the existence could detect the agent.

1 はじめに

本研究では、会話によるコミュニケーションとエージェントの人間らしい振る舞いに着目する。人間と会話を行う人工知能の一例として、人工無脳（会話ロボット）がある[1]。人工無脳は、人間が発した言葉の部分的な文言に対して、パターンマッチングであらかじめ用意された文を返すことで会話を行っているように見せかけることができる。有名な人工無能の一つに ELIZA [2]がある。ELIZA は精神科医代行システムであり、特定の用途においては驚くほどうまく会話を成立させてみせた。しかし、パターンマッチングによる会話には限界があり、実社会における複雑な会話をうまく成立させることは未だ困難である。

一方、会話によるコミュニケーションを主題としたゲームとして人狼ゲームが注目されている[3]。人狼ゲームは、情報の駆け引きや高度な推論が必要となるパーティゲームである。篠田らは汎用人工知能の標準問題として、人狼ゲームを検討している[4]。人工知能研究の標準問題としては、これまでチェス、将棋、囲碁、追跡問題、囚人のジレンマ、RoboCup などさまざまな問題が提案されてきた。人狼ゲームはこれらの標準問題と比較して、会話によるコミュニケーションでゲームが進行する、不完全情報ゲーム

である、非決定性（合議制）を有する、他者の思考を推論する、他者を説得するという点などが特徴的であるとされる。また、本研究では、人間に扮するという人狼ゲームの特徴にも着目する。エージェントが人間に適切に擬態することで、抵抗感なく人間社会に溶け込むことができると考えられる。

人狼ゲームを題材とした研究は始まったばかりであり、未知な部分が多い。これまで強化学習を用いた強いエージェントを目指す研究[5]や実際の人間のプレイログから行動モデルを構築し人間らしいエージェントを目指す研究[6]などが行われている。

本研究では、人間とエージェントによる人狼ゲームのプレイログを分析し、人間に擬態することでエージェントと見破られない人間らしいエージェントの実現を目指す。本稿では、エージェントの存在を秘匿した場合、仄めかした場合、および明かした場合の3つの条件で人狼ゲーム実験を行うことで、人間に擬態するエージェントの要素を検討する。

2 人狼ゲーム

2.1 人狼ゲーム

人狼ゲームは、プレイヤー同士の駆け引きをコンセプトとしたコミュニケーションパーティゲームである。本ゲームのカバーストーリーを以下に示す。

とある平和な村に、人の見た目をした狼（人狼）が紛れ込みます。人狼は夜になると村人の誰か1人

*連絡先:岡山大学大学院自然科学研究科
〒700-8530 岡山県岡山市北区津島中 3-1-1
E-mail: takata.k@mif.sys.okayama-u.ac.jp

を食い殺してしまいます。昼間は村人が全員起きているので、さすがの人狼も多人数には勝てないためおとなしくしています。この昼間の時間で、村人たちは村に紛れ込んだ人狼を探しだして処刑します。しかし人狼は人の見た目をしているので、誰が人狼か村人にはわかりません。村には村人と人狼以外に、人狼かどうか見分ける能力をもった占い師や、人狼の味方をする狂人など、様々な能力、特徴をもった人がいます。村人たちは彼らの話す情報を元に誰が人狼かを暴きだして、村から人狼を排除するため毎日一人ずつ処刑していきます。村が全滅してしまうのが先か、人狼を処刑して平和が訪れるのが先か、村人達の生存を賭けた戦いが今はじまる！[7]

本ゲームは、会話によるコミュニケーションや情報の不完全性、他者の説得など実社会の多くの要素を有している。プレイヤーが一カ所に集まって行う対面型ゲームが広く行われているが、インターネットの普及に伴い掲示板などを利用した BBS 型ゲームも行われるようになった。人狼ゲームと一言に言っても様々な国、形態で行われているため、数多くのローカルルールが存在する。そのため、本研究で対象とする人狼ゲームのルールを次節に示す。

2.2 ゲームルール

本研究では、プレイヤー数はゲームマスターを除く 8 人とし、BBS 型人狼ゲームを用いる。

ゲーム開始時に各プレイヤーに役職が割り振られ、役職に応じて村人陣営と人狼陣営に分かれて各陣営の勝敗を競う。役職配分は、村人 4 人、占い師 1 人、霊媒師 1 人、人狼 2 人とする。各役職の説明を表 1 に示す。役職によって情報の多寡が異なっており、例えば人狼プレイヤーは仲間の人狼プレイヤーを知ることができるが、その他のプレイヤーは自分以外の役職が伏せられている。村では、人狼によって毎日村人陣営の中から 1 人のプレイヤーが襲撃され、ゲームから除外される。また、生存プレイヤー全員の投票によって毎日 1 人のプレイヤーが処刑され、ゲームから除外される。村人陣営は人狼を村からすべて処刑する、人狼陣営は人間の数を人狼以下にすることが勝利条件である。

ゲームは、昼ターンと夜ターンが交互に繰り返されることで進行する。昼ターンでは人狼を探し出すために占い師の占い報告や処刑者の投票に関する村全体の話し合いが行われ、同時に村人陣営に隠れて人狼同士による襲撃先の相談などが行われる。そして、昼ターンの終わりに処刑者の投票や占い先の選定、襲撃先の投票が行われる。昼ターンは 8 分の制限時間がある。夜ターンでは処刑の執行、能力の使用、襲撃などが順に行われ翌日の昼ターンへ移動す

表 1 本実験における人狼ゲームの役職とその説明

| Role | Group | Given abilities |
|----------|----------|---|
| Villager | Villager | No special abilities |
| Seer | Villager | Knowing the group of a designated participant in nights |
| Medium | Villager | Knowing the group of ever executed participants in nights |
| Werewolf | Werewolf | Knowing the other werewolves from scratch and talking with the others as telepathic communication Attacking to one of villager group every night |

る。なお、1 日目は占いのみが行われ、処刑や襲撃は行われない。また処刑と襲撃では毎日必ず 1 人処刑および襲撃される。

3 実験概要

3.1 実験目的

選択回答式の人狼ゲームにおいて、人間に擬態するエージェントに求められる要素を抽出する。

3.2 選択回答式の人狼ゲーム

本実験では、従来の人狼ゲームのようにプレイヤーが自然言語を用いて自由に会話を行いながらゲームが進行するのではなく、発言の選択肢を選択することによって会話を行いながらゲームを進行する。選択回答式に会話を行うことで、本来日本語に含まれる語尾や表現のゆらぎによる他者の印象変化を排除し、人間に擬態するための要素を行動様式に焦点を絞って分析することができると考えられる。発言の選択肢は、人狼知能プロジェクト[8]で提案されている会話プロトコルに従う。

また、エージェントを交えて選択回答式の人狼ゲームを行うために、人狼知能プロジェクトで提供されている図 1 のような対戦用インタフェースを用いて人狼ゲームを行う。実験参加者は、本インタフェースを通して他プレイヤーとやりとりする。

表 2 に本実験で用いた村全体の会話における発言テンプレートを示す。発言の順番が回ってきた参加者は、プルダウンメニューから発言の種類を選択したのち、表 2 の鍵括弧部分を別のプルダウンメニューから選択することで発言する。話すことがないときは発言の順番をスキップすることができる。秘密裏に行われる人狼同士の会話も同様に行う。また、本インタフェースでは、ゲーム開始時に各参加者に

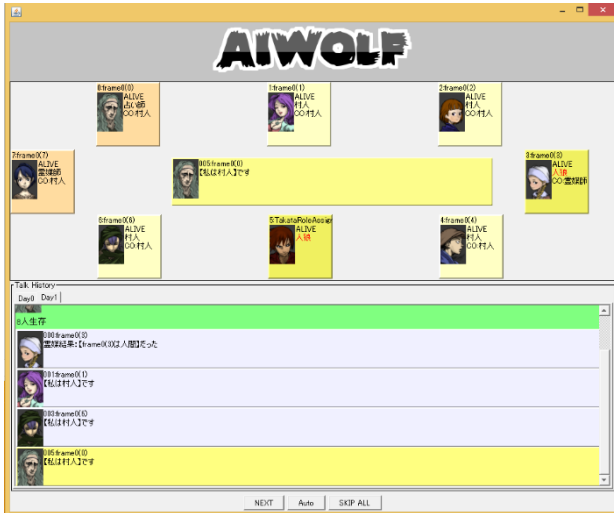


図1 対戦用インタフェース

表2 村全体の会話における発言テンプレート

| Types of statement | Chat template |
|--------------------|---|
| Vote | 「Character」に投票する |
| Coming-out | 私は「Role」です |
| Estimate | 「Character」は「Role」だと思う |
| Divine | 占い結果: 「Character」は「Werewolf or Human」だった |
| Inquest | 霊媒結果: 「Character」は「Werewolf or Human」だった |
| Agree | >> 「Day」: 「Order」に同意 (The content of agreement) |
| Disagree | >> 「Day」: 「Order」に反対 (The content of disagreement) |

表3 各実験におけるエージェントの条件

| # of exp. | The exposure of agent existence | Games that an agent played | | | |
|-----------|---------------------------------|----------------------------|-----|-----|-----|
| | | 1st | 2nd | 3rd | 4th |
| 1 | Hidden | ○ | ○ | ○ | ○ |
| 2 | Hinted | - | ○ | - | ○ |
| 3 | Known | ○ | ○ | ○ | - |

ランダムでキャラクターが割り振られるため、他の参加者がどのキャラクターでプレイするかわからないようになっている。

3.3 実験条件

実験参加者は20代男性7人2組と8人1組の計22人とし、表3に示すように各組に対して異なる条

件の実験（以下、実験1、実験2、実験3とする）を行う。すべての実験に共通して、人間8人あるいは人間7人とエージェント1体の8プレイヤーで4ゲーム行う。本実験ではエージェントは必ず人狼の役職が割り振られる。

実験1では7人1組の参加者を対象とし、4ゲームすべてエージェントが参加する。ただし、参加者に扮した実験協力者1人がエージェントのダミー役として参加し、見かけ上は人間8人でゲームをプレイする。

実験2では8人1組の参加者を対象とし、4ゲームのうち第2ゲームと第4ゲームにエージェントが参加する。ただし、エージェントが参加するゲーム回においては、参加者の中でエージェントを割り振られた参加者がダミー役を演じる。

実験3では7人1組の参加者を対象とし、4ゲームのうち第1、2、3ゲームにエージェントが参加する。第4ゲームはエージェントの代わりに実験者が参加するが、見かけ上は人間7人とエージェント1体でゲームをプレイする。

エージェントの参加について、実験ごとに秘匿度合いを変更する。実験1は参加者全員に対してエージェントが参加していることを秘匿し、全ゲーム終了後にエージェントの存在を明かす。実験2は実験の最初に1体のエージェントの存在を仄めかし、全ゲーム終了後にエージェントの参加したゲーム回を明かす。実験3は実験の最初にエージェントが1体参加することを通知し、全ゲーム終了後に第4ゲームにおいてエージェントが参加していないことを明かす。

3.4 収集データ

すべての実験に共通して、実験参加者の役職、勝敗、会話ログ、戦略、誰が人狼であるかの推測（以下、人狼の推測）、主観評価、および誰がエージェントであるかの推測（以下、エージェントの推測）を収集した。役職、勝敗、会話ログはシステムログから収集し、主観評価はアンケートによって収集した。アンケートは、人狼ゲームの経験に関する2項目（初回のみ）と自分のプレイに対する自己評価や他者のプレイに対する評価など村人陣営の参加者は8項目（初回10項目）、人狼陣営の参加者は7項目（初回9項目）に5段階評価で回答させた。

戦略は、参加者にゲームプレイ中にどのようにふるまうかの方針を自由に記述させることで収集した。ゲーム開始時（ゲーム内の初日）に基本戦略を記述し、ゲームプレイ中に戦略を変更したい場合は変更戦略を、追加したい場合は追加戦略をゲーム中に随時記述させた。

人狼の推測は、村人陣営の役職を割り振られた参加者の自由記述により収集した。ゲーム内の毎昼ターン中に記述させた。

エージェントの推測は、情報開示状態に合わせて記述方法を変更した。実験 1 では、全ゲーム終了後に参加者に第 3, 4 ゲームに 1 体のエージェントが存在したことを明かし、エージェントを特定させた。第 1, 2 ゲームについては日を跨ぐため、回答させなかった。実験 2 では、各ゲーム終了時にアンケートと合わせて、エージェントの有無を回答させた。エージェントがいたと回答した場合は誰がエージェントであったかの推測を記述させた。実験 3 では、各ゲーム終了時にアンケートと合わせて、誰がエージェントであったかの推測を記述させた。なお、推測を行う際に該当ゲームの会話ログを確認させた。

3.5 実験の流れ

すべての実験に共通して、以下の (1), (2) をゲーム開始直前に 1 回通して行ったのち (3), (4) を 4 回繰り返した。

(1) 実験概要の説明

参加者全員に実験概要の説明を行った。説明内容として、実験の流れや対戦用インタフェースの操作方法について説明した。また、ゲーム中に行ってもらう戦略と人狼の推測、エージェントの推測（実験 1 を除く）の記述方法を説明した。

(2) 人狼ゲーム練習プレイ

参加者に各自の PC で対戦用インタフェースを立ち上げ TCP/IP 通信でクライアントとしてサーバー PC に接続させ、実際にゲームプレイを進めながら操作の確認などを行わせた。

(3) 人狼ゲーム本プレイ

本プレイでは、各ゲーム開始時に参加者全員に事前にこちらで設定した役職を割り振った。設定方法としては、エージェントには必ず人狼の役職を割り振り、他の参加者には 4 回の本プレイのうち占い師か霊媒師か人狼の役職を必ず 1 回から 2 回割り振るように設定した。ゲームプレイ中は、対戦用インタフェースを用いてゲームを進めた。また、戦略と人狼の推測をエクセルファイル等に記述させた。村人陣営あるいは人狼陣営が勝利すると本プレイ終了とした。

(4) アンケート

参加者に 1 プレイ毎にそのゲーム回におけるアンケートに回答させた。また、合わせてエージェントの推測を記述させた。ただし、実験 1 については、全ゲーム終了後に参加者全員に 1 体のエージェントの存在を明かし、第 3 ゲームおよび第 4 ゲームにおけるエージェントの推測をまとめて記述させた。

4 実験結果と考察

本稿では、収集データの中からエージェントの推測と会話ログに焦点を絞り、人間に擬態するエージェントの要素を検討した。

表 4 にゲームの勝敗結果を示す。実験 1 については村人陣営が第 1, 3, 4 ゲームで勝利、人狼陣営は第 2 ゲームのみ勝利であった。実験 2 については 4 ゲームすべて人狼陣営が勝利し、実験 3 については 4 ゲームすべて村人陣営が勝利した。

表 5, 表 6, 表 7 に実験 1, 実験 2, 実験 3 におけるエージェントの推測の結果を示す。各ゲーム回（実験 1 については第 3, 4 ゲームのみ）においてエージェントと推測されたプレイヤー名を示している。なお、実験 2（表 5）においてエージェントがいないと推測された場合は”Nobody”とした。また、表 6 において、エージェントの代わりにゲームに参加した実験者をエージェントと推測した場合は”Pz”とした。さらに、エージェントを見抜いた回答とエージェントが不参加であることを見抜いた回答については括弧書きで”Right”を加えた。

次に、エージェントの推測の結果をもとに各実験においてエージェントが見破られた割合（以下、看破率）を表 8 に示す。実験 1 においてエージェントの看破率は 1/14（約 7.1%）であった。実験 2 においてエージェントの参加した第 2, 4 ゲームにおけるエージェントの看破率は 4/14（約 28.8%）であった。また、全ゲーム回において”Right”にみられるように正しい回答をした割合は 12/30(40.0%)であった。実験 3 においてエージェントの参加した第 1, 2, 3 ゲームにおけるエージェントの看破率は 14/21（約 66.7%）であった。

表 8 上段のエージェントの看破率についてみていく。参加者が自分以外の 7 人のプレイヤーからランダムにエージェントを選出した場合にエージェントを的中させる割合の期待値は 1/7（約 14.3%）である。ただし、実験 2 ではエージェントが参加しているかどうか判断しているため、実験 2 の期待値は 1/8（12.5%）として考える。これら理論値と本実験における看破率を比較すると、実験 1 の第 3, 4 ゲームおよび実験 2 の第 2 ゲームにおいては、ほぼ同程度であることからエージェントがほとんど見破られることなくプレイできていたと考えられる。一方、実験 2 の第 4 ゲームおよび実験 3 全体においては理論値を大きく上回っていたため、エージェントが看破されていたと考えられる。すなわち、エージェントの存在が疑われる条件下では、エージェントの看破率は高まっていた。

表 4 ゲームの勝敗

| # of exp. | Winners | | | |
|-----------|---------|-----|-----|-----|
| | 1st | 2nd | 3rd | 4th |
| 1 | V | W | V | V |
| 2 | W | W | W | W |
| 3 | V | V | V | V |

※ V: villager, W: werewolf

表 5 エージェントの推測 (実験 1)

| Participants | Player's estimations | |
|--------------|----------------------|-----|
| | 1st | 2nd |
| Pa1 | Pf1 | Pg1 |
| Pb1 | Agent (Right) | Pf1 |
| Pc1 | Pd1 | Pe1 |
| Pd1 | Pf1 | Pb1 |
| Pe1 | Pc1 | Pd1 |
| Pf1 | Pb1 | Pb1 |
| Pg1 | Pc1 | Pf1 |

表 6 エージェントの推測 (実験 2)

| Participants | Player's estimations | |
|--------------|----------------------|---------------|
| | 1st | 2nd |
| Pa2 | Nobody (Right) | Pg2 |
| Pb2 | Pe2 | Nobody |
| Pc2 | Ph2 | |
| Pd2 | Pf2 | Pe2 |
| Pe2 | Pa2 | Nobody |
| Pf2 | Nobody (Right) | Agent (Right) |
| Pg2 | Pa2 | Nobody |
| Ph2 | Nobody (Right) | Nobody |
| Participants | 3rd | 4th |
| Pa2 | Pc2 | Agent (Right) |
| Pb2 | Pa2 | Agent (Right) |
| Pc2 | Nobody (Right) | Nobody |
| Pd2 | Nobody (Right) | |
| Pe2 | Nobody (Right) | Agent (Right) |
| Pf2 | Nobody (Right) | Nobody |
| Pg2 | Nobody (Right) | Nobody |
| Ph2 | Pa2 | Pg2 |

表 9 に各実験において表 2 で示した発言の種類毎に村全体の会話における参加者の発言数 (エージェントと実験者の発言は除外) を示す。なお, "Num." は発言数, "Per." はその実験の総発言数に対する割合を示している。

表 7 エージェントの推測 (実験 3)

| Participants | Player's estimations | |
|--------------|----------------------|---------------|
| | 1st | 2nd |
| Pa3 | Agent (Right) | Agent (Right) |
| Pb3 | Pa3 | Agent (Right) |
| Pc3 | Agent (Right)t | Agent (Right) |
| Pd3 | Pg3 | Pg3 |
| Pe3 | Agent (Right) | Agent (Right) |
| Pf3 | Agent (Right) | Agent (Right) |
| Pg3 | Agent (Right) | Pb3 |
| Participants | 3rd | 4th |
| Pa3 | Agent (Right) | Pz |
| Pb3 | Agent (Right) | Pz |
| Pc3 | Pb3 | Pg3 |
| Pd3 | Agent (Right) | Pz |
| Pe3 | Agent (Right) | Pz |
| Pf3 | Pg3 | Pz |
| Pg3 | Pf3 | Pz |

表 8 各実験におけるエージェントの看破率

| # of exp. | Percentage of correct answers when an agent played | | | | |
|-----------|--|-------|-------|-------|-------|
| | 1st | 2nd | 3rd | 4th | Ave. |
| 1 | - | - | 14.3% | 0.0% | 7.1% |
| 2 | - | 14.3% | - | 42.9% | 28.8% |
| 3 | 71.4% | 71.4% | 57.1% | - | 66.7% |
| # of exp. | Percentage of correct answers | | | | |
| | 1st | 2nd | 3rd | 4th | Ave. |
| 1 | - | - | 14.3% | 0.0% | 7.1% |
| 2 | 37.5% | 14.3% | 62.5% | 42.9% | 40.0% |
| 3 | 71.4% | 71.4% | 57.1% | 0.0% | 50.0% |

各実験における参加者の発言数について, 実験 1 は推測発言, 実験 2 および実験 3 はカミングアウト発言が多い傾向にあり, 各実験の集団によって発言の傾向に差が生じていた可能性がある。今回実験 1 の第 1, 2 ゲームを除いて, エージェントは最初に処刑されゲームから脱落しているため, ゲーム序盤の動向に着目する。表 10 にゲーム内の 2 日目 (最初の処刑や襲撃が行われる日) までの参加者 (人間) の発言数を示す。表 10 と表 9 を比較すると, 2 日目までの発言の種類について, 実験 1 の発言数の割合に大きな変化はないが, 実験 2 および実験 3 はカミングアウト発言の比重が非常に大きくなっている。実験 1 では参加者はエージェントの存在を知らないため, ゲームに勝利することだけを考えてプレイして

いたと考えられる。一方、実験2および実験3では参加者はエージェントの存在を疑っているため、ゲームの勝敗を競うだけでなく、エージェントを探すまたはあぶり出す行動を意識的あるいは無意識のうちに行っていたことが考えられる。このような心理状態ではうかつな発言を行うことを避けて周囲の発言に合わせる傾向にあり、特にゲーム序盤では発言の種類が同じものに偏ったのではないかと考えられる。すなわち、エージェントの存在が疑われる条件下では発言の傾向が偏るのではないかと考えられる。しかし、本実験では参加者がエージェントを探すまたはあぶり出す行動を行ったかどうか計測していないため、追実験等により検証する必要がある。

次に、表11に各実験におけるゲーム内の2日目までのエージェントの発言数を示す。ただし、エージェントの推測を収集していない実験1の第1,2ゲームにおける発言数は除外した。表11と表10を比較すると、実験2および実験3において、エージェントは比較的多様な発言をする傾向にあるが、参加者はカミングアウト発言に偏っている。すなわち、エージェントと参加者で発言の傾向が異なっていたために、エージェントの看破率が高くなったのではないかと考えられる。

また、ゲーム中のエージェントの具体的な行動をみていく。実験2および実験3では、ほとんどの参加者が割り振られた役職に関係なく、ゲーム序盤に「私は村人です」というカミングアウト発言を行っていたのに対して、エージェントは占い師や霊媒師を騙るとき以外にカミングアウト発言を行ってなかった。また、参加者からエージェントに対して人狼の疑いをかけられた場合にエージェントが反応を示さない、あるいは無関係な発言をする場面が見受けられた。すなわち、発言の傾向が異なるために、場の流れに適応していないエージェントの行動が目立ったと考えられる。適応的でない行動は、人狼ゲームにおける他のプレイヤーの不信感の増加につながる一般的な要素の一つであり、人間らしくない要素の一つでもあると考えられる。したがって、エージェントと参加者で発言の傾向が異なっていたために、エージェントの適応的でない行動が目立ち、エージェントの看破率が高くなったのではないかと考えられる。

以上のことから、エージェントの存在が疑われる条件下では、参加者の発言が偏る傾向にあり、その偏りに適応していないエージェントの行動が目立され、エージェントの看破率が高くなったと考えられる。すなわち、エージェントの存在が疑われる条件下では、エージェントの存在が秘匿されている場合より適応的な行動をとる必要があると考えられる。

表9 村全体の会話における実験参加者の発言数

| Types of statement | Exp. 1 | | Exp. 2 | | Exp. 3 | |
|--------------------|--------|------|--------|------|--------|------|
| | Num. | Per. | Num. | Per. | Num. | Per. |
| Vote | 22 | 0.14 | 9 | 0.08 | 7 | 0.05 |
| Coming-out | 31 | 0.20 | 39 | 0.36 | 71 | 0.52 |
| Estimate | 64 | 0.41 | 25 | 0.23 | 17 | 0.13 |
| Divine | 6 | 0.03 | 3 | 0.03 | 5 | 0.04 |
| Inquest | 3 | 0.02 | 5 | 0.05 | 4 | 0.03 |
| Agree | 19 | 0.12 | 15 | 0.14 | 22 | 0.16 |
| Disagree | 12 | 0.08 | 11 | 0.10 | 10 | 0.07 |
| Total | 157 | 1.00 | 107 | 1.00 | 136 | 1.00 |

表10 2日目までの実験参加者の発言数

| Types of statement | Exp. 1 | | Exp. 2 | | Exp. 3 | |
|--------------------|--------|------|--------|------|--------|------|
| | Num. | Per. | Num. | Per. | Num. | Per. |
| Vote | 8 | 0.10 | 3 | 0.06 | 0 | 0.00 |
| Coming-out | 21 | 0.26 | 32 | 0.63 | 51 | 0.82 |
| Estimate | 31 | 0.39 | 6 | 0.12 | 2 | 0.03 |
| Divine | 4 | 0.05 | 1 | 0.02 | 2 | 0.03 |
| Inquest | 0 | 0.00 | 0 | 0.00 | 1 | 0.02 |
| Agree | 9 | 0.11 | 8 | 0.16 | 6 | 0.10 |
| Disagree | 7 | 0.09 | 1 | 0.02 | 0 | 0.00 |
| Total | 80 | 1.00 | 51 | 1.00 | 62 | 1.00 |

表11 2日目までのエージェントの発言数

| Types of statement | Exp. 1 | | Exp. 2 | | Exp. 3 | |
|--------------------|--------|------|--------|------|--------|------|
| | Num. | Per. | Num. | Per. | Num. | Per. |
| Vote | 0 | 0.00 | 1 | 1.00 | 1 | 0.33 |
| Coming-out | 2 | 0.33 | 0 | 0.00 | 1 | 0.33 |
| Estimate | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Divine | 2 | 0.33 | 0 | 0.00 | 1 | 0.33 |
| Inquest | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Agree | 2 | 0.33 | 0 | 0.00 | 0 | 0.00 |
| Disagree | 0 | 0.00 | 0 | 0.00 | 0 | 0.00 |
| Total | 6 | 1.00 | 1 | 1.00 | 3 | 1.00 |

5 おわりに

本研究では、人狼ゲームにおいて人間に擬態することでエージェントと見破られない人間らしいエージェントの実現を目指している。本稿では、選択回答式の人狼ゲームにおいて、エージェントの存在を秘匿した場合、仄めかした場合、および明かした場合の3つの条件で人間とエージェントによる人狼ゲームを行い、人間に擬態するエージェントに求められる要素について分析を行った。その結果、エージェントの存在が疑われる条件下では、ゲーム内の発言が偏る傾向にあり、その偏りに適応していないエージェントの行動が注目され、エージェントの看破率が高くなった可能性が示唆された。したがって、エージェントの存在が疑われる条件下では、エージェントの存在が秘匿されている場合より適応的な行動をとる必要があると考えられる。今後の課題として、参加者によるエージェントを探すまたはあぶり出す行動の有無の検証、人狼以外の役職を演じるエージェントを使用した実験、エージェントが人間に擬態することによるゲームの勝敗への影響の分析等を行っていく予定である。

謝辞

本研究に際しまして、実験の準備や進行の補助に協力いただいた岡山大学工学部機械システム系学科の高木智也氏に感謝いたします。

参考文献

- [1] 富坂亮太, 鈴木崇史: 人工無脳 (会話ロボット), 映像情報メディア学会誌, Vol. 64, No. 1, pp. 64-66, (2010)
- [2] Weizenbaum Joseph: ELIZA—a computer program for the study of natural language communication between man and machine, Communications of the ACM, Vol. 9, No. 1, pp. 36-45, (1966)
- [3] 片上大輔, 鳥海不二夫, 大澤博隆, 稲葉通将, 篠田孝祐, 松原仁: 人狼知能プロジェクト, 人工知能学会誌, Vol. 30, No. 1, pp. 65-73, (2015)
- [4] 篠田孝祐, 鳥海不二夫, 片上大輔, 大澤博隆, 稲葉通将: 汎用人工知能の標準問題としての人狼ゲーム, 人工知能学会全国大会論文集, No. 28, pp. 1-3, (2014)
- [5] 梶原健吾, 鳥海不二夫, 大澤博隆, 片上大輔, 稲葉通将, 篠田孝祐, 西野順二, 大橋弘忠: 強化学習を用いた人狼における最適戦略の抽出, 情報処理学会第76回全国大会講演論文集, pp. 597-598, (2014)
- [6] 平田佑也, 稲葉通将, 高橋健一, 鳥海不二夫, 大澤博隆, 片上大輔, 篠田孝祐: プレイログから獲得した行動選択確率を用いた人狼ゲームのシミュレーション, 人工知能学会全国大会論文集, No. 29, pp. 1-4, (2015)
- [7] 人狼道: 初心者でもできる人狼入門, <http://jin-rodou.com>, (2015.10.16 参照)
- [8] 鳥海不二夫, 稲葉通将, 大澤博隆, 片上大輔, 篠田孝祐, 梶原健吾, 松原仁, 狩野芳伸: Artificial Intelligence based Werewolf | Let's play "Werewolf" with AI!, <http://www.aiwolf.org/>, (2015.10.16 参照)