人狼知能サーバの構築

鳥海 不二夫[1,a)]　梶原 健吾[1]　大澤 博隆[2]　稲葉 通将[3]　片上 大輔[4]　篠田 孝祐[5]

概要：人工知能を用いたゲームをプレイするエージェントは数多く開発されているが，現在までに，人工知能が人間に勝利しているテーブルゲームの多くは全ての情報が公開されている完全情報ゲームである．それに対して，ゲームの中には情報が完全には公開されておらず，情報の被均一性がゲーム性を演出する不完全情報ゲームや，ゲームの本質がプレイヤ同士の自由対話や交渉によって実現されるコミュニケーションゲームがある．本研究では，不完全情報コミュニケーションゲームである人狼ゲームをエージェントがプレイするサーバを構築し，人とエージェント，エージェントとエージェントがそれぞれゲームをプレイする環境を整える．

Development of AI Wolf Server

Toriumi Fujio[1,a)]　Kajiwara Kengo[1]　Osawa Hirotaka[2]　Inaba Michimasa[3]　Katagami Daisuke[4]
Shinoda Kosuke[5]

Abstract: There are many Artificial Intelligent Agents which are trying to play the games, however, there are no agents which can play communication game with humans. In this paper, we developed the game server to play "Are you a werewolf?", which is one of most famous communication game in the world, by artificial intelligent agents.

1. 背景

人工知能を用いたゲームをプレイするエージェントは数多く開発されている．特に，テーブルゲームにおいては研究が盛んに行われており，1997 年にはチェスのグランドチャンピオンに，2013 年には将棋のプロ棋士にそれぞれ勝利するエージェントが作成されている．

現在までに，人工知能が人間に勝利しているテーブルゲームの多くは全ての情報が公開されている完全情報ゲームである．また，ゲームが盤面上でのみ行われるため，コンピュータ上でゲームをモデル化し，再現することが容易である．

それに対して，ゲームの中には情報が完全には公開されておらず，情報の被均一性がゲーム性を演出する，不完全情報ゲームが存在する．さらに，ゲームの本質がプレイヤ同士の自由対話や交渉によって実現されるコミュニケーションゲームがある．

人狼ゲームはコミュニケーションのみにより勝敗が決定する，人間の持つ極めて高度な認知能力を駆使して行う不完全情報型のコミュニケーションゲームである．将棋や囲碁といった完全情報ゲームとは異なり多くの情報がプレイヤーによって隠蔽される．各プレイヤーは会話と行動から隠された情報を推測しつつ，自らの秘密は隠蔽したままチームの勝利に向けて発言，行動していく．人狼ゲームには，プレイヤーが持つ情報の非対称性，信頼を得る説得・協調行動，嘘を見抜く推論など従来の人工知能分野では扱っていなかった多数の解決すべき問題が存在する．

人狼に関する研究としては，人狼ゲームの数学的考察 [1], [2] やプレイログのデータ分析 [3], [4] は存在するが，人狼をプレイするエージェントを実現した研究はない．

[1] 東京大学
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113–8656, Japan
[2] 筑波大学 University of Tsukuba
[3] 広島市立大学 Hiroshima City University
[4] 東京工芸大学 Tokyo Polytechnic University
[5] 電子通信大学 The University of Electro-Communications
[a)] tori@sys.t.u-tokyo.ac.jp

本研究の最終目的は，その人狼ゲームを人間の代わりにプレイできるエージェント (人狼知能) の実現にある [5], [6]．このようなエージェントは，他者との対話により状況を判断する能力，他者を説得する能力，他者を騙す能力などを有する必要性がある．特に，多段階の自己認識，すなわち「自分は A である」「自分は A であると B は思っている」「自分が A であると B が思っていることを，B は知っているだろう」といった相手から見た自分の状態を認識することは，これまでの人工知能の分野ではあまり扱われておらず，今後の人工知能の発展に大きく寄与する効果が期待できる．さらには，人間と対戦するエージェントであればゲームとして面白くするため適度に騙されることや，より「人間らしい」説得などを実現する必要がある．

本研究では，エージェントが人狼ゲームを実現するサーバを構築し，人とエージェント，エージェントとエージェントがそれぞれゲームをプレイする環境を整える．その上で，プロトコル及びサーバを公開し，多くの研究者が参加可能な人狼エージェントの大会を実施する．これによって，様々なバックグラウンドを持った研究者による集合知の獲得を目指す．

2. 人狼ゲームの概要

2.1 人狼の概要

人狼ゲームは，アメリカのゲームメーカー Loony Labs. が 2001 年に発売されたパーティーゲーム「汝は人狼なりや」及びその派生ゲームの総称である．多数の類似ゲームが世界中で市販され，世界中でプレイされている．日本においてもタブラの狼 (2) やうそつき人狼 (3) など多数のゲームが販売されている．人狼をプレイする方法としては，前述したような市販のカードなどを使って行う対面型と，WEB 上のアプリケーションを使って行う BBS タイプが存在する．日本における BBS タイプの人狼の内，最も盛んにプレイが行われているサービスの一つが人狼 BBS(4) である．人狼 BBS ではこれまでに数千回以上のゲームが行われており，ログデータを用いた研究 [3][4] も行われている．

将来的には対面プレイが可能な物理エージェントを構築することが本プロジェクトの目的の一つとなっているが，現時点では難しい．そこで，当面は BBS タイプの人狼を対象に考え，基本ルールを人狼 BBS に準拠させることとする．それに伴い，人狼 BBS におけるプレイログをエージェントの設計などに利用することが可能となる．

2.2 ゲームの流れ

プレイヤーにはまずランダムに「役職」が割り当てられる．プレイヤーは役職によって，人間または人狼陣営にそれぞれ振り分けられ，各プレイヤーはチームの勝利を目指す．人間陣営の目標は人狼の全滅に，人狼陣営の目標は人間の人数を人狼の人数と同数以下にすることにあり，目標を達成した陣営の勝利となる．各自の役職は本人以外には非公開であるため，自分以外の誰がどの役職か分からない．特に，人間側は誰が人狼か分からないため，会話の中から人狼を探し出すことが基本的な行動指針となる．一方，人狼陣営のプレイヤーは同じ人狼陣営のプレイヤーをゲーム開始時に知らされる．そのため，人狼陣営に所属するプレイヤーは互いに協力しながら，人間陣営に正体がばれないように行動することが基本的な行動指針となる．

ゲームは昼と夜の 2 つのフェーズからなる．昼のフェーズでは全てのプレイヤーによって，誰が人狼かを探し出すための議論が行われる．このとき，後述する各種能力を持った役職についているプレイヤーは当該能力によって知り得た情報を用いて，自分たちの陣営が有利になるように議論を導くことになる．一定期間の議論の後，プレイヤー全員の投票によって，人狼と考えられる人物を処刑する．処刑されたプレイヤーはゲームから除外され，ゲーム終了まで参加することが出来ない．

夜のフェーズでは，人狼陣営に所属するプレイヤーが人間陣営のプレイヤーを一人選び，襲撃する．襲撃されたプレイヤーは死亡者として扱われ，処刑されたプレイヤーと同様にゲームから除外される．また，各種能力を持った役職には，夜のフェーズに能力に応じた情報を与えられる．昼のフェーズと夜のフェーズを繰り返し，勝利陣営を決定する．

議論において，人間陣営に所属するプレイヤーは人狼の嘘を見破るかが最大のポイントとなる．また，能力を持つ役職に就いたプレイヤーは能力によって知り得た情報を使って他のプレイヤーを説得することがポイントとなる．一方，人狼陣営のプレイヤーは自分たちが不利にならないように議論を誘導し，時には能力を持った役職であると偽り，議論を間違った方向へ誘導することなどが基本プレイとなる．

2.3 人狼ゲームにおける主な役職

人狼には様々なバージョンが存在し，役職もバージョンによって異なるが，人狼 BBS に準拠する本プロジェクトでは以下の役職がいるものとする．

(1) 村人
　人間陣営に所属する．特に何の能力も持たない．

(2) 占い師 (予言者)
　人間陣営に所属する．夜のフェーズで占い結果として指定した一人が「人狼であるか否か」を知ることが出来る．人間陣営においては最も重要な役職である．

(3) 霊媒師
　人間陣営に所属する．前日に追放した人物が人狼かどうかを知ることが出来る．

(4) 狩人 (ボディーガード)

人間陣営に所属する．夜のフェーズでプレイヤー１人を指定して，人狼の襲撃から守ることが出来る．狩人が守ろうとしたプレイヤーを人狼が襲撃した場合，その日は誰も死なないことになる．

(5) 共有者 (フリーメーソン)
人間陣営に所属する．二人一組の役職で，ゲーム開始前にもう一人の共有者が誰かを知ることが出来る．

(6) 人狼
人狼陣営に所属する．複数人狼がいる場合は，互いにコミュニケーションを取ることが可能である (対面の人狼の場合，夜のフェーズで目配せによるコミュニケーションを取る)．夜のフェーズで任意の村人を襲撃することが可能．

(7) 狂人
人狼陣営に所属する．ただし，人狼からは誰が狂人であるかは把握されず，能力も無い．村人と同様であるが，人狼陣営が勝利したときに勝利となるため，ひそかに人狼に協力をする．勝利人数のカウント時には人間陣営として数えるため，狂人が自ら処刑される事によって勝利することも可能である．

2.4 人狼ゲームの基本戦略

2.4.1 人間陣営の基本戦略

人間陣営の基本戦略は人狼陣営の嘘を見抜くことにある．

人狼を探し出すには占い師と霊媒師による情報が重要である．しかしながら，多くの場合人狼陣営のプレイヤーが偽の占い師および霊媒師として名乗り出る (以下 CO: Coming Out) するため，誰が真の占い師，霊媒師であるかを見抜くことが必要となる．

たとえば，二人の占い師の内一人が自分を人狼であると指摘した場合，当該プレイヤーは人狼陣営に所属していることが分かる．このような情報を積み重ねていくことによって，各プレイヤーは誰が人狼かを絞り込んでいく．

また，人間陣営では占い師や霊媒師がどのタイミングで CO するかを全員の相談であらかじめ決めておき，人狼陣営のプレイヤーが人間陣営のプレイヤーを騙す要素が少なくなるようにプレイすることが多い．

2.4.2 人間陣営役職持ちの基本戦略

人間陣営で役職を持つプレイヤーは占い師，霊媒師，狩人，共有者である．

このうち，占い師は出来るだけ早く人狼を見つけ出すことが必要である．しかしながら，もし人狼陣営に占い師であることが発覚すると襲撃の対象となるため，占い師であることは隠しておくことも多い．ただし，人狼陣営のプレイヤー (人狼または狂人) が「自分が占い師である」と嘘を付くことがあるため，その場合は自分が本物であると信頼を勝ち取るため他のプレイヤーを説得することが必要となる．

霊媒師は，処刑されたプレイヤーが人狼であるかどうかを判断できるが，能動的なアクションは起こしづらいため，多くの場合人狼が処刑されたときに名乗り出てその事実を告げることになる．ただし，人狼陣営のプレイヤーが霊媒師を騙った場合は対抗して名乗り出る必要もある．

狩人は他のプレイヤーを襲撃から守ることが出来る重要な役職であるが，自分自身を守ることはできないため，自ら狩人であることを名乗ることは少ない．狩人であるとばれないように真の占い師を守ることが重要である．また，襲撃から誰かを守った場合，守られたプレイヤーが人間陣営であることが分かるため，他のプレイヤーよりも情報が多くなる場合もある．その場合は，得られた情報を使って議論をうまく誘導する必要がある．

共有者は，お互いが人間であることが分かっているため強い信頼関係を維持することができる．また，それぞれが相方を指定しながら名乗り出ることで，他のプレイヤーに人間側であることを強く説得することができるため，人間陣営で信頼できる人物としてプレイすることが可能である．また，万が一名乗り出る前に偽の占い師に人狼であると判定されれば，もう一方の共有者にも偽の占い師が判明するため，大きな情報となる．

2.4.3 人狼陣営の基本戦略

人狼陣営の基本戦略は人間陣営を騙して人狼陣営のプレイヤーが処刑されないように議論を誘導することにある．

人狼陣営のプレイヤーは基本的に人間陣営にいるように振る舞う．ただし，占い師の情報によって正体が暴かれる危険があるため，占い師を見抜き襲撃する，あるいは占い結果を他のプレイヤーが信じないよう議論を誘導する必要がある．

そこで，多くの場合複数いる人狼陣営のプレイヤーの内何人かが占い師や霊媒師であると名乗り出る戦略 (騙り) を採用することが多い．このとき，人狼同士で話し合って決める場合もあれば，狂人が勝手に名乗り出る場合もある．

人狼が占い師を騙った場合，人狼とそれ以外のプレイヤーが分かっているため，常に正しい占いを行うことが出来るため，疑われづらい．一方狂人が占い師を騙った場合は誰がどの役職かは分からないため，間違った占いを行ってしまう可能性もあるが，それによって処刑されても人間の人数が減るため，人狼陣営の勝利に貢献することになる．

3. 人狼知能エージェントの設計

3.1 人狼知能サーバの構成

人狼ゲームをエージェントによってプレイさせるためのプラットフォームとして，人狼ゲームサーバを構築した．人狼ゲームサーバはサーバとクライアントに分かれている．サーバはゲームマスターの役割と試合ログの保存，通信の制御を行なう．人狼サーバとクライアントは TCP/IP もしくはシステム内部 API によって通信を行なう．これ

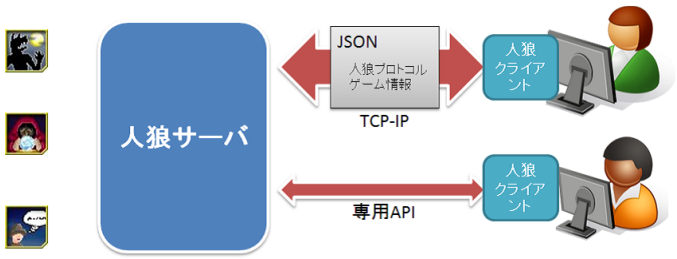

図 1　人狼サーバ

によりネットワーク越しの対戦及び，システム内での高速シミュレーション実験の双方を実現する (Figure.1)．

各ユーザが構築するエージェントはクライアントへ接続して，ゲーム状態の情報収集及び行動の宣言を行う．各クライアントに接続するエージェントは参加者が独自に開発可能であり，人工知能によるエージェントの構築はもちろん，WizardOfOZ 形式の人間が操作するエージェントを導入することも想定される．ゲームサーバおよびクライアントは Java を用いて作成されている．

TCP-IP 通信でやりとりされる情報は JSON を用いているため，JSON を理解するクライアントであれば，サーバへ接続可能である．今後は多様な言語によってエージェントが構築可能となるよう，環境を整備しているところであり，現在 C によるクライアントを作成中である．

3.2　サーバ・エージェント通信

エージェントとサーバをつなぐ通信はサーバからのリクエストにエージェントが応える形で行われる．

エージェントから送られるリクエストを表 1 に示す．エージェントがこれらのリクエストをサーバから受信した場合，リクエストに応じてサーバに応答を返す．なお，このとき同時にエージェントが知りうる情報がサーバからエージェントに送られる．エージェントはこれらの情報を処理して，リクエストにどのように応えるを決定する．

サーバから送られる JSON の例を表 2 に示す．送られてくる JSON には大きく gameInfo と request があり，gameInfo には現在のゲームの状況が示されている．

gameInfo の各変数が情報が何を示すかについては，表 3 に示す．例に挙げた情報からは，

- 自分はエージェント番号 0
- 現在 1 日目．(0 日目からスタート)
- エージェント 0，2 は生存，エージェント 1 は死亡
- エージェント 0 と 5 が人狼
- 昨日処刑されたのはエージェント 6
- 昨日襲撃されたのはエージェント 1

などの情報を獲得することができる．なお，ここで襲撃投票先 (attackVoteList) や囁き [*1](whisperList) は役職が人狼であるエージェントにしか提供されない．同様に，占い結果 (divineResult) なども対応する役職のエージェントにのみ提供される．

[*1] 人狼同士の相談

表 1　Requests from Server

Request	内容
Initialize	ゲーム開始時の初期化
DailyInitialize	一日の開始時の初期化
Finish	ゲーム終了
Name	エージェントの名前
Role	希望する役職
Talk	発話
Whisper	人狼同士の発話 (囁き)
Vote	投票先指定
Divine	占い先指定
Guard	護衛先指定
Attack	襲撃先指定

表 3　GameInfo

変数名	内容
agent	このエージェントの固有番号
day	何日目か
statusMap	各エージェントの状況
roleMap	既知の役職
executedAgent	昨日処刑されたエージェント
attackedAgent	昨日襲撃されたエージェント
divineResult	占いの結果 (占い師のみ)
mediumResult	霊媒の結果 (霊媒師のみ)
voteList	昨日の投票
attackVoteList	襲撃投票 (人狼のみ)
talkList	この日の対話ログ
whisperList	この日の囁きログ (人狼のみ)

また，サーバからのリクエストは request に格納されており，ここでは Talk が要求されているため，エージェントは人狼プロトコルに従って発話文をサーバに返送することになる．

3.3　人狼エージェントの作成

人狼エージェントは前述した通りサーバと TCP/IP 通信を行い，サーバから送信される情報を理解できればよい．しかし，それだけでは作成が困難であるため，エージェント作成ライブラリを用意している．

現在エージェント作成ライブラリには Java で作成したものが用意されている．エージェント作成ライブラリには，サーバと TCP/IP 通信を行うためのライブラリとエージェントを作るためのライブラリが含まれている．

エージェント作成者は，エージェントの動作を決定する Player インターフェースを実装したクラスを作成し，TCP/IP 通信を行うクライアントクラスである TcpIpClient のインスタンスを用いてサーバへと接続する．本ライブラリを用いることで，エージェント開発者は通信部分を考え

表 2　サーバから送られるリクエストの例


```
{
  ”gameInfo”:{
    ”agent”:0,
    ”day”:1,
    ”statusMap”:{”0”:”alive”, ”1”:”dead”,”2”:”alive”,···},
    ”roleMap”:{”0”:”werewolf”,”5”:”werewolf”},
    ”executedAgent”:6,
    ”attackedAgent”:1,
    ”divineResult”:null,
    ”mediumResult”:null,
    ”voteList”:[
      {”agent”:0,”day”:0,”target”:11},
      {”agent”:1,”day”:0,”target”:11},
      {”agent”:2,”day”:0,”target”:6},
      ⋮
    ],
    ”attackVoteList”:[
      {”agent”:5,”day”:0,”target”:1},
      {”agent”:0,”day”:0,”target”:9}
    ],
    ”talkList”:[
      {”agent”:4,”content”:”Agent[03] werewolf”,”day”:1,”idx”:0},
      {”agent”:3,”content”:”Agent[00] werewolf”,”day”:1,”idx”:1},
      ⋮
    ],
    ”whisperList”:[]
  },
  ”request”:”Talk”
}
```


ることなく，サーバから送られてくる情報とリクエストに基づいて，どのように返答すればよいかのみを実装すればよい．たとえば，処刑者を投票時する際には vote メソッドが呼ばれるため，それまでに得られたゲーム情報に基づいて投票したいエージェントを決定し，当該エージェントのインスタンスを返せばよい．また，各種情報クラスが用意されており，サーバから送られてくる GameInfo クラスのインスタンスを通して取得することができるため，データ形式 (JSON) を意識する必要はない．

なお，それとは別にエージェントの開発を補助するライブラリも用意されており，人狼プロトコルを意識せずにプログラミングするための人狼プロトコルライブラリも用意されている．

3.4　人狼サーバにおける人狼ゲームの流れ

人狼サーバを用いた人狼ゲームの流れを以下に示す．

(1) 初期化・エージェントの接続 (希望役職の送信)
(2) エージェントの役職決定
(3) 一日の開始
(4) 各エージェントによる発言
(5) 人狼による囁き
(6) 全エージェントが発言のおよび囁きを終了していなければ (4) に戻る
(7) 投票・占い・護衛・襲撃先の決定
(8) 投票・占い・護衛．襲撃の処理．
(9) 勝敗が決まっていなければ，一日進め (3) に戻る
(10) 人狼の生存数が 0 ならば人間側の勝利，人狼と人間が同数ならば人狼側の勝利

以上に従って，人狼エージェントは一日ごとに会話，(囁き，) 投票，(特殊能力の適用) を繰り返していく．その中で，他のエージェントの発話や投票行動から人狼を推定し処刑する，あるいは他のエージェントに人狼であることを悟られないように襲撃を繰り返すことによって勝利を目指す．

3.5　人狼プロトコル

人狼は自然言語によって行われるコミュニケーションゲームであるが，人工知能に自然言語によるゲームを行わせるためには課題が多い．そこで，人狼専用の対話プロトコルを開発し，エージェント同士のコミュニケーションを

表 4　人狼プロトコルの例

プロトコル表現	意味
Agent[00] comingout medium	私 (Agent[00]) は霊媒師です
Agent[06] medium_telled HUMAN	霊媒の結果、 Agent[06] は人間だった
Agent[05] werewolf	Agent[05] は人狼だ (と思う)

実現する．

現在の人狼ゲームサーバでは，簡易プロトコルセットによってコミュニケーションを行うためのライブラリを提供している．簡易プロトコルセットでは，複雑な類推を説明することは出来ず，自分の「職業のカミングアウト」,「自分が知っている情報の提供」,「怪しいと考えているエージェント」のみを発話することが可能である．簡易プロトコルセットを用いた発話の例を表 4 に示す．

なお，人狼プロトコルの設計については，文献 [5] に詳細が記載されている．

4. 終わりに

現在人狼サーバは人狼知能プロジェクトページ (http://aiwolf.org) において公開中である．また，GitHub(https://github.com/aiwolf/) でもサーバアプリケーションを公開中であり，開発への協力者を募集中である．今後は，2014 年 11 月に人狼知能作成のためのチュートリアルを開催し，2015 年 3 月ごろに簡易プロトコルを用いた人狼エージェントによる人狼大会を行う予定である．その参加者も現在募集中である．

謝辞　本研究を進めるに当たり，様々な助言と補助をいただいた松原仁先生に感謝いたします．

参考文献

[1] Piotr Migdał. A mathematical model of the mafia game. *arXiv preprint arXiv:1009.1031*, 2010.

[2] Erlin Yao. A theoretical study of mafia games. *arXiv preprint arXiv:0804.0071*, 2008.

[3] 稲葉通将, 大畠菜央実, 鳥海不二夫, 高橋健一. 雑談ばかりしてると殺される-人狼 bbs におけるプレイヤーの発言傾向と意思決定・勝敗の分析-. *JAWS 2013*, 2013.

[4] 稲葉通将, 鳥海不二夫, 高橋健一. 人狼ゲームデータの統計的分析. ゲームプログラミングワークショップ 2012 論文集, 2012.

[5] 大澤博隆. コミュニケーションゲーム「人狼」におけるエージェント同士の会話プロトコルのモデル化. HAI シンポジウム, 2013.

[6] 鳥海不二夫, 稲葉通将, 大澤博隆, 片上大輔, 篠田孝祐, 西野順二. 人工知能は人狼の夢を見るか－人狼知能プロジェクト－. 日本デジタルゲーム学会, 2014.