第 24 回インテリジェント・システム・シンポジウム [2014.9.18－19]

FAN-14-016

人工知能標準問題としての人狼ゲームの提案

A proposal of Ai Werewolf Game Challenge for Standard AI Problem

○ 篠田孝祐 (電通大)　　鳥海不二夫（ 東大）　　稲葉通将（ 広市立）
大澤博隆（ 筑波大）　　片上大輔（ 東京工芸大）

Kosuke Shinoda, The University of Electro-Communications, kosuke.shinoda@uec.ac.jp
Fujio Toriumi, The University of Tokyo
Masamichi Inaba, Hiroshima City University
Hirotaka Osawa, University of Tsukuba
Daisuke Katagami, Tokyo Polytechnic University

We propose a standard problem “ Are you a Werewolf? ” for artificial intelligence. This game is one of communication game around the table. We has been thought this game is useful metrics for evaluating progress artificial intelligence.

Key Words: Werewolf game, Standard AI Problem, AI Werewolf Project

1　はじめに

「 ある村に， 人間の姿に化けられる人喰い人狼が現れた. 人狼は人間と同じ姿をしており， 昼間には区別がつかず, 夜になると村人たちを一人ずつ襲っ ていく． 村人たちは疑心暗鬼になりながら， 話し合いによっ て人狼と思われる人物を１人ずつ処刑していくことにした」・・・ 以上が， コミュニケーショ ンゲームとして知られる人狼ゲームのカバーストーリーである． この人狼と呼ばれるゲームがある． このゲームは， 基本プレイヤ同士の話し合いにより進行し勝敗がきまる． この人狼ゲームは， 言語のみで進行するゲームでありながら， 全世界で楽しまれているゲームであり， 特に日本ではオンライン上で行う BBS 人狼が日々 行われている． 本論文では， この人狼ゲームを， 人工知能記述の評価のための標準問題として用いることを提案し， その可能性を議論する.

2　人工知能における代表的標準問題

標準問題とは， ある技術の性能・ 発展などを評価するうえで， 提案された手法やモデルを比較可能とするために設ける問題のことである． 人工知能が対象とする諸問題は, 非常に幅広く多様である． そのため， 異なる研究者で， 同じ領域を対象とした課題を設けても， 必ずしも同様の環境を構築できるとは限らず， 得られた知見の再現性に乏しいために単純に比較することは難しい． そこで， 互いがそれぞれの問題を各々 手段で実装して調べる代わりに， 現実におけるさまざまな課題の中で重要かつ一般的な問題を十分に表現でき現実の問題へのアプローチとなりうる問題を, 標準問題とすることで人工知能技術の進歩発展を評価することが試みられてきた.

これまで， 人工知能の標準問題としてさまざまな問題が提案されてきた． 代表的なものとしては， チェ ス [1] や将棋・ 囲碁， 追跡問題， 囚人のジレンマ， RoboCup[2, 3] などがある． チェ ス・ 将棋・ 囲碁などはゲーム進行のスコアリングと探索の問題であり， 追跡問題は共通の全体ゴールをもつ集団が全体の戦略と個々 の戦術との関係を追及する問題である． そして， RoboCup では， ソフトウェ アとしての人工知能技術とロボッ トというハードウェ アを融合しサッ カー・ 災害救助・ 日常環境という実社会への投入を前提として問題を提案している． これらに共通するのは， それぞれの課題に対して， 共有するべき大きな目標を設定し, 共通のフレームワーク上で互いの技術・ 知見を実装することで， 利点欠点を比較可能となる.

3　人狼知能の提案

3.1　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 [4] 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やうそつき人狼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も行われている [5, 6].

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．

このように人狼ゲームは， 不完全情報化で会話によって各プレイヤが認識している情報のエントロピーを操作するコミュニケーションゲームの一種である． 我々は， この人狼ゲームを， 人もしくはエージェント間でのプレイすることを目標とした「 人狼知能プロジェクト」 を立ち上げた．

3.2 人狼ゲームの特性

3.2.1 コミュニケーションによるゲーム進行

これまでのゲームの多く， 特に完全情報ゲームでは， 環境の状況（ 他のエージェントの振舞も含む） をもとに行動を決定するため， 状況の評価とそれを含めた探索効率の向上がゲームの主要素であった． それに対して， 人狼ゲームでは， そのアクションだけでなくコミュニケーションによる情報交換によりゲームを進行し勝敗を決する．

鳥海ら [7] が， ゲームの状況をもとにプレイヤの行動を学習することでゲーム展開に有意な影響を与える行動を抽出できることを示したように， 会話による要素がゲーム展開を支配している分けではない． だが， エージェント同士の対戦だけでなく， 人間も交えてゲームを行うことを検討した場合には， すべてのプレイヤが合理的に行動するとは限らず， それと同時にすべての参加者が合理的に行動するとは信じられないために， コミュニケーションが個々のプレイヤがもつ情報や考えの確からしさに影響を与えると考える． また， コミュニケーションは， 人とゲームを楽しむ要素として非常に重要でもある．

3.2.2 情報の不完全性と非決定性

将棋や囲碁のような完全ターン性のゲームでは， 互いの状況を完全に把握できる環境であるのに対して， 人狼ではお互いの情報をある程度秘匿することでゲームが成立する．また， 決定すべき行動に投票などあるため， 個々の行動のみで場の状況が決定することはなく非決定的でもある．

3.2.3 信頼の構築• 説得

情報の不確定性をベースにしているが， そこから情報を確定させていくためには， 自己の持つ情報を開示するなどして説得する作業が求められる． ただし， プレイヤには複数の陣営があるため， 他の陣営から味方の陣営の情報のエントロピーを意識して活動する必要がある． そのためにも，信頼を獲得し説得する作業が重要である． そのため， 他者の行動目的や保有情報などの推定が必要であり， さらに他者からみた自己意図がどのようなものであるのかを推定することが必要である． これらに合わせて， 情報開示などを行うことで， 他者から自身への信頼を構築するかどうかは，このゲームの特徴の一つである．

3.2.4 第三者的視点からの情報非決定性

人間同士の行うゲーム， 特にボードゲームは， いくつかの種類に分かれている． その中で「 コミュニケーションゲーム」 と総称される分野がある． コミュニケーションゲームではゲームの勝敗がプレイヤー同士の情報交換に依存する．コミュニケーションゲームの中で広く知られた例としては，プレイヤー同士の交渉が勝敗を決定するモノポリー[1] やカタン[2] がある． モノポリーやカタンでは， プレイヤー同士のコミュニケーションによる説得や協調により利得が変化する． 人狼は， このようなゲームの中でもっとも極端な形を持つゲームであり，「 コミュニケーション」 以外の客観的情報， 勝敗決定要因がほとんど存在しない． 人狼で客観的と言える情報は， 各人の発言内容自体と， 日数， 処刑者，襲撃者などであり， 人狼に登場する殆の役職達は， 自分自身の役職をゲームに参加していない第三者にも「 客観的に」証明する手段を持たない．

具体例を挙げると， あなたが「 村人（ 無能力者）」 であり，あなたの友人がそのゲームを外から観察している（ ＝ゲーム上の発話のみを観測している） とする． この時， ゲームプレイ中に， あなたが村人， あるいは人間側であることをその友人に対し証明することは不可能である（ これは， あなたがどの役職であっても同様である）． もちろん， ゲーム中の「 人狼」 のプレイヤーにとってあなたが少なくとも「 非人狼＝人間側」 であることは自明であるし， もし村の中の占い師があなたを占ったとすると， その占い師はあなたが人間側であることを知ることができ， また占い師はあなたが人間であるという発話を行うことができる． しかしながら， 村の外にいるあなたの友人から見た時， その「 人狼」 のプレイヤーが本当に人狼であるのか， あるいは「 占い師」 と宣言した人間が本当に占い師であるのかどうかはゲーム終了までわからない． 人狼というゲームにおいて，情報はもちろん， 他者の目的も共有できる状況は限られている． このような， 客観的視点からの情報決定ができないのは人狼ゲームの基本原則であり， プレイヤはそれを前提として戦略・ 戦術を決めていく必要がある．

3.3 人狼知能の目指すところ

人狼知能プロジェクトでは， 人狼知能を人工知能の標準問題とするに当たって， ゲーム環境では以下の段階を検討している．

第 1 段階: ソフトウェアエージェントにより構成された村での人狼プレイの実現

第 2 段階: ソフトウェアエージェントと人間が混合した村での人狼プレイの実現

第 3 段階: ロボットエージェントと人間が混合した村での人狼プレイの実現

また， それぞれにおいて会話の自由度の段階も人狼のゲーム性に大きな影響をあたえるために人狼知能でも， 以下の段階を考慮する．

第 A 段階: 会話に用いる単語ならびに論理記述を限定した対話環境

[1] http://ja.wikipedia.org/wiki/モノポリー
[2] http://ja.wikipedia.org/wiki/カタンの開拓者たち

第 B 段階: 自然言語による対話環境

第 C 段階: 音声と限定された身体表現による対話環境

第 D 段階: 自由な音声と身体表現による対話環境

A 段階は, 自然言語を容易に処理できる環境が整うまでに相当の時間が必要ではないかと考えられるために, 現実的にはさらに細かい条件を設けて緩和する方法を模索することになる. また, C,D 段階とは身体性に関わる表現であり、 こちらも言語と同様に制限された環境から自由な表現へと拡張することを想定している。

人狼ゲームを構成する主たる要素は他にも考え得るが, まずは上記の段階をもとに人狼知能の研究環境を構築する. まずは, 1A の段階として, ソフトウェアエージェント同士が人狼ゲームを行うための会話プロトコルである人狼プロトコル [8] を用いて対戦が可能となる人狼サーバ [7] のプロトタイプを実装した.

3.4 人狼ゲームサーバ

人狼サーバは, ゲームにおけるゲームマスターと試合ログの保存ならびに通信の制御を行う. 人狼サーバを用いてゲーム参加するエージェントならびに人は, 人狼プロトコルを通してエージェント間の会話ならびにゲーム進行や情報をやりとりする. 人狼サーバの詳細に関しては, [7] や人狼知能プロジェクト [3] のページを参考されたい.

4 人狼知能にもとめられる人工知能技術

人狼知能を実現する過程において様々な人工知能技術が求めら得ると考えられる. 主なものとして下記の要素は最低限必要となる.

- 記憶・ プランニング・ 推論・ 学習

  この要素は, 人狼ゲームに限らず既存ゲームでも必要とされている研究課題であり, 人狼でも変わらず必要である.

- 他のプレイヤへの同調・ 反駁・ 説得

  これも既存の組織的活動が求められる標準問題でも自己・ 他者モデルは必要な技術である. 人狼を含めた不完全情報下で行われる多人数ゲームでの特有な要素として, 同調や他者同士の競合をつくりだすためには, 他者にある自己を含めた他者のモデルを推測できるモデルを必要とする.

- コミュニケーション・ インターフェス

  通常我々の行う人狼ゲームは自然言語を用いて行うが, 現時点においてソフトウェアエージェントなどを交えてゲームを行うには自然言語で行うことは難しい. そこで, まずはゲームプロトコルを設定するなどすることで, ゲーム性を損なわない範囲で限定された言語を用いてゲームを行う. さらに, 実際のゲームでは人の行動や表情がゲームの展開に影響を与えてしまうこともある. そのため, それらもまずは限定された範囲で表情などの振る舞いを加えてゲームを行うことを見当する必要がある.

[3] http://www.aiwolf.org/

5 人狼知能における人工知能研究課題

人狼知能における研究課題の特徴は, コアとして, 不完全情報下における他者の心理の理解と会話を通した場の情報エントロピーの操作があり, ゲームの基盤を構築するための課題として, ゲームプロトコルの開発や人間とロボットの対話や自然言語処理などがある. 以下, いくつか具体的な研究課題を示す.

5.1 ゲーム性を損なわない会話プロトコルの設定

人狼ゲームを初めコミュニケーションゲームでは, 会話は間違いなく重要な要素である. しかしながら, 自然言語を初めから扱うには現段階においては困難である. そのため, ゲーム性を損なわいが, ゲームを行うには十分な対話プロトコルの設計が本プロジェクトの初期段階では重要である. また, 将来的に自然言語によるゲームを行う状況までに, どのような要素を緩和していくことでゲーム性を大きく変質させないでゲームを行えるのかも, プロジェクトの展開には重要な課題である.

5.2 人狼コーパスからの戦略獲得

現時点では, 人狼 BBS から得られたゲームデータから, 各プレイヤの行動パターンなどから目的・ 意図を推定し戦略を抽出する課題がある. 今後は, 実際の人狼ゲームのプレイ動画などから人狼コーパスを生成することができるようになることで, BBS 人狼と対面人狼とのゲーム性・ 戦略戦術の違いなども分析出きるようになる.

5.3 人狼における定石の発見

人狼ゲームには, セオリーと呼ばれるヒューリスティックな常套手段はあるが, 定石と呼べるようなゲーム戦術は, 現時点では明らかになっていないと考える. これは, 客観的なゲーム状況の把握が難しいためゲームの展開を完全に記述が困難であり, 状況を具体的に評価するための指標がないためだと考える. したがって, ゲーム状況を客観的に評価して定石を記述できる

5.4 HAI: Human Agent Interaction

まずは, ソフトウェアエージェントを, 将来的にはロボットなど身体をもったエージェントによって人狼知能をプレイすることを想定している. その際, エージェントを作成するということはもちろん、 人間のグループで行われる人狼ゲームとエージェントを交えて行うそれとにどのような違いがあるのかなど, HAI に関連する研究課題は多々ある. また, 会話が重要な要素である人狼では, 自然言語によるコミュニケーションを成立すると同時に意図推定なども研究が必要な課題であり、 人狼はそのテスト環境として適切であると考える.

5.5 ゲームにおける同調・引き込み: 楽しさを作り出す構造

人狼ゲームは, 日本に限らず世界各地でおこなわれているゲームである. そのゲームが盛り上がる構造, 参加者がそれぞれ楽しめるための要件などを明らかにすることで, ゲームによる教育や訓練における引き込み要素を検討できる.

5.6 人間力教育ツールとしての有用性の検証

これらコミュニケーションゲームを活用することで, 学生の人間教育に有用であるという示唆を得ている. また, 人狼ゲームを入社時の評価手段の一つとして導入している企業もあると聞く. したがって, 人狼ゲームを知育ゲームとして活用するのに有用な可能性も高いと考える. しかし

ながら， 人狼のどのような要素が， 人の能力を引き出すのかや伸ばすのかは分かっていない． そのため， ゲームがプレイヤに与える心理的な影響を検討する必要もある．

6 汎用的な知性の評価の手法

近年， 三度人工知能が注目を浴びており， それと重なるように汎用人工知能 (Generic Artificial Intteligence) が注目を集め始めている [9]． 汎用人工知能とは， 人間レベルの知能を実現することを目標とした人工知能研究のことである． この汎用人工知能の具体的な姿は人工知能や人間の汎用知能と同じく明確な定義は存在しないが， Adams らにより， 汎用人工知能の実現にむけた課題や評価シナリオが整理されている [Adams2011]． 汎用人工知能に関する詳細な説明は， 本稿では割愛するが， Adams らは， 汎用人工知能の評価のためのシナリオとして以下の 6 つのシナリオを提案している．

1. General Video-game Learning

   特定のビデオゲームを対象にするのでなく， さまざまビデオゲームをプレイしながら学びクリアーできるようにするという課題． 主に視覚処理と入力操作の課題を内包したシナリオ

2. Preschool Learning

   基本的な運動能力や絵を理解する能力を育てるシナリオ

3. Reading Comprehension

   読解力をみにつけるためのシナリオ

4. Story/Scene Comprehension

   文章を理解したうえで， 場面における登場人物の関係性や物語の流れを読み解く力を学ぶシナリオ

5. School Learning

   Preschool と関連あるシナリオではあるが， 論理的な思考や社会的な関係性などをまなぶことを想定したシナリオ

6. The Wozniak Test

   見知らぬ家を訪問しその家の主にコーヒーを給仕することを目標としたシナリオ． ある意味 RoboCup Home の課題と近い目標であるが， 社会関係や他者の理解まで要求しているシナリオ

Adams らは， これらのシナリオをベースとして， 必要とする環境とそこで行うべきタスクを設定することを求めている． また， 彼らは， AGI の実現に必要とされる能力の領域も提案している． 具体的には， 知覚・ 記憶． 注意・ 社会インタラクション・ プランニング・ モチベーション・ 推論・ 運動作業・ コミュニケーション・ 学習・ 情動・ 自己/他者モデル・ 構築/創造・ 定量化など 15 の能力エリアを設けている．

人狼ゲームをソフトウェアエージェントにより対戦するゲームとしたうえで， 人工知能という観点から見たときに， それを実現するには， エージェントやインタラクション・ 認知科学・ 心理学， ロボット工学など多様な領域にまたがる． Adams らの挙げた能力エリアでいうと， ソフトウェアエージェントとして実装するだけでも， 記憶・ 社会的インタラクション・ プランニング・ 推論， コミュニケーション・ 学習・ 自己/他者モデル・ 定量化など多岐にわたる能力が必要となる． つまり， 人狼ゲームは， 汎用人工知能を評価するための標準問題としてもある程度カバーできる可能性がある．

7 まとめと今後の課題

本論文では， 人工知能のための標準問題として人狼ゲームの可能性を検討した． 今後はこのゲームをベースとした競技会などを設けることにより， 普及を図る予定である． そして， 集合知的アプローチにより人狼知能の実現を目標とする．

なお， 我々の人狼プロジェクトにご興味を持たれた方は， プロジェクト HP（ http://www.aiwolf.org/） をご覧いただければ幸いである．

References

[1] Murray Campbell, “Knowledge discovery in deep blue”， Communications of the ACM, vol.42-11, pp.65–67, 1999.

[2] Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda, Eiichi Osawa, “RoboCup: The Robot World Cup Initiative”, AGENTS’97, pp.340-347, 1997

[3] Hiroaki Kitano, Minoru Asada, Yasuo Kuniyoshi, Itsuki Noda, Eiichi Osawa, Hitoshi Matsubara, “RoboCup: A challenge problem for AI”, AI Magagin, Vol.18, No.1, pp.74-85, 1997

[4] Wikipedia: Werewolf(game) `http://en.wikipedia.org/wiki/Werewolf_(game)`

[5] 稲葉通将, 鳥海不二夫, 高橋健一 (2012) “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, vol. 2012, no. 6, pp.144147

[6] 稲葉通将, 大畠菜央実, 鳥海不二夫, 高橋健一 (2013) “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-”, JAWS2013， 2013

[7] 鳥海不二夫，梶原健吾，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? ー人狼知能プロジェクトー”， 日本デジタルゲーム学会， 2014

[8] 大澤博隆，鳥海不二夫，片上大輔，篠田孝祐，稲葉通将 “人狼ゲームのプロトコル設計:推理と説得のプロコトル”， FAN2014， 2014 (発表予定)

[9] am S. Adams et al.: “Mapping the Landscape of Human-Level Artificial General Intelligence”， AAAI， 2011.