

OS-14

人狼知能と不完全情報ゲーム

AI Werewolf and Incomplete Information Games

稲葉 通将
Michimasa Inaba
電気通信大学
The University of Electro-Communications.
m-inaba@uec.ac.jp

片上 大輔
Daisuke Katagami
東京工芸大学
Tokyo Polytechnic University.
katagami@cs.t-kougei.ac.jp

狩野 芳伸
Yoshinobu Kano
静岡大学
Shizuoka University.
kano@inf.shizuoka.ac.jp

大槻 恭士
Takashi Otsuki
山形大学
Yamagata University.
otsuki@yz.yamagata-u.ac.jp

Keywords: werewolf game, incomplete information game, multi-agent simulation, natural language processing, human computer interaction.

1. はじめに

2016年、完全情報ゲームの中でも人間に勝つことは困難であるといわれていた囲碁において、トッププロがコンピュータに敗北したことから、人工知能が対象とするゲームは完全情報ゲームよりも難しいゲームに軸足を移すことになった。そこで近年注目が集まっているのは、ポーカーや麻雀に代表される不完全情報ゲームと、さらにそれにコミュニケーションの要素が加わった不完全情

報コミュニケーションゲームである。本オーガナイズドセッションは、不完全情報コミュニケーションゲームである「汝は人狼なりや？」(以下、人狼ゲームと表記)を対象として研究を行っている「人狼知能プロジェクト」のメンバを中心に、ゲーム情報学だけでなくエージェント、インタラクション、自然言語処理などさまざまなセッションで散発的に発表が行われていた人狼ゲーム研究、および不完全情報ゲーム研究を一つのセッション内で議論するとともに、研究者間の交流を図るために企画したものである。



図1 JSAI 2019 オーガナイズドセッション「人狼知能と不完全情報ゲーム」の様子

2. 人狼知能研究

人狼知能研究は、競技 (Competition) の側面と、研究 (Research) の側面の両側面があり、両者が互いに影響を与えつつ発展することにより、さまざまな知見が蓄積されるものと考えられる。これらをうまく運営し、両者のコラボレーションを促進することが、人狼知能プロジェクトの一つの目標である。

競技 (Competition) の側面においては、これまでにさまざまな競技のためのセミナーや競技大会を企画してきた。競技大会としては、第1回人狼知能大会が、ゲーム開発者の会議 CEDEC 2014@ パシフィコ横浜で8月

に開催された。この大会を皮切りに、CEDEC 2018 に至るまで4回の人狼知能大会が開催されている。そして今年8月、人工知能に関する国際会議IJCAI 2019@マカオの交渉エージェントの競技セッションANAC Competitionにおいて、記念すべき第1回人狼知能大会の国際大会が開催された。これを皮切りに国際的に大きく発展していくことが期待される。

研究 (Research) の側面においては、これまでに研究ワークショップにおけるイベントの企画や、人狼知能プロジェクトの各メンバによる研究活動、そして昨年の2018年度人工知能学会全国大会(第32回)@鹿児島市城山観光ホテルにおいて、オーガナイズドセッション「人狼知能と不完全情報ゲーム」を企画し、8件の研究発表と、特別企画人狼知能プレ大会2018を開催した。そして、今年6月新潟市朱鷺メッセ新潟コンベンションセンターにおける2019年度人工知能学会全国大会(第33回)において、2回目となるオーガナイズドセッション「人狼知能と不完全情報ゲーム」が開催された。今年度も研究の内容として、会話のモデル化、統計分析、行動分析、データ構築、役職・陣営推定、ノンバーバル動作の分析などさまざまな研究発表が行われた。会場の様子を図1に示す。

本稿では、上記の2019年度人工知能学会全国大会において開催されたオーガナイズドセッション「人狼知能と不完全情報ゲーム」の内容を中心に詳しく報告する。

3. 発表内容の概観

2019年度の本OSに投稿された研究は9件であり、そのうち7件が人狼ゲームに関するものであった。人狼に関する研究を大きく「人間どうしの人狼ゲームにおけるプレーヤの分析」、「人狼ゲームの戦略的分析」、「人狼ゲームにおける役職の推定」の三つに分類し、発表された研究の概要を紹介する。

3.1 人間どうしの人狼ゲームにおけるプレーヤの分析

人間どうしで行う人狼ゲームのプレーヤの分析は、人狼知能プロジェクトの発足以前のものも含め、いくつか研究例が存在する。特に海外では人狼が嘘をつくという側面に着目し、欺瞞検知を行うため、プレーヤそれぞれの話の長さや回数、話を遮った回数などを特徴として用いた研究 [Chittaranjan 10] や、手や頭の動きを用いた研究 [Xia 07]、発話の長さを用いた研究 [Zhou 08] などがある。最近では、欺瞞検知にとどまらず、広くプレーヤの振舞いを分析する研究が活発に行われており、本OSで発表された2件もジェスチャと生体情報という側面から分析を行っている。

まず、筑波大学の高山周太郎らによる「実世界人狼ゲームのジェスチャの分析」[高山 19]では、5名のプレーヤが人狼ゲームをプレイする様子を撮影し、プレイ中の

ジェスチャとゲームの結果の関連性について分析した。分析の結果、狂人と占い師の腕の開き具合がゲームの勝敗に関連していることを明らかにした。腕の開き具合はその場のコミュニケーションにおけるリーダーシップとの関係性が指摘されており、各陣営で重要な役職である狂人と占い師がリーダーシップを発揮できるかがゲームの勝敗に影響することを意味している。

京都産業大学の山本浩隆らによる「生体信号に基づく人狼プレイヤーの「強さ」分析の試み」[山本 19]では、同じく5名の人狼ゲームを対象とし、ゲーム中において役職カミングアウトや投票など重要なイベントが発生した際の皮膚電気活動の振舞いについて、勝率の高いプレーヤと低いプレーヤとで比較を行った結果について報告した。分析の結果、勝率の高いプレーヤは重要なイベントの際に皮膚電気活動の上昇が見られたが、勝率が低いプレーヤの場合、重要なイベントでかつそのプレーヤが何らかの行動を行った場合にのみ皮膚電気活動の上昇が見られた。このことは、強いプレーヤは自分が当事者でなかったとしても、重要なイベントを戦略的に注視していることが示唆されたと報告している。

3.2 人狼ゲームにおける役職の推定

強い人狼知能を実現することは、人狼ゲーム研究の目的の一つである。人狼知能プロジェクトが主催する人狼知能大会のうち、プロトコル部門においては勝率で優勝者を決定している。また、近年は深層学習を用いた人狼知能の設計に関する研究も行われている [大川 17, 王天 18]。

本OSでは、2件の役職推定に関する発表が行われた。人狼知能において役職の推定は極めて重要であり、特に人狼を精度良く発見できることはその強さに直結する。

東京都市大学の福田宗理らによる「人狼ゲームにおける会話情報による役職推定」[福田 19]では、大川貴聖らによって提案された役職推定のためのニューラルネットワークモデルを拡張し、より推定精度の高いモデルを構築した。具体的には、大川らのモデルは役職推定対象のプレーヤがどのような発言や行動を行ったかということの特徴としていたが、本研究では、推定対象のプレーヤ以外の発言や行動も考慮するモデルとした。

日本大学の小村友希らは「人狼ゲームにおける明示的役職・陣営推定理由の抽出」[小村 19]と題して、役職・陣営推定のための解釈可能な推定モデルの抽出を行った。推定モデルにはランダムフォレストと決定木を使用し、小村らが定義した約90種の説明変数を用いた。説明変数は、占い・霊能結果、他のエージェントへの肯定発言・否定発言の回数などである。また、ランダムフォレストにより作成された推定モデルからのルール抽出のため、inTreesを用いた。実験の結果、解釈可能なモデルルールを得ることができたが、モデル自体の推定精度が高いとはいえず、また、推定不可能な役職が存在した

ことから、ルールの妥当性に関しても疑問が残る結果となったことを報告した。

3.3 人狼ゲームの戦略の分析

人狼ゲーム自体の性質や特徴を明らかにする研究も活発に行われている。人狼ゲームではさまざまなルールや役職が存在し、ルールと役職の組合せによってもゲーム性が大きく変化するため、その解析は非常に難しい。そこで、これまでに役職を人狼と村人のみに限定したルールでのゲームの数学的な解析を行った研究 [Migdal 10, Yao 08] や、最適戦略を求める研究 [Braverman 08] などが行われてきた。人狼と村人のみのルールでは、追放プレーヤと襲撃プレーヤはランダムに決定するとみなすことができるため、その解析は比較的容易であった。一方、大澤博隆らはより複雑なルールとして、村人、占い師、人狼の3プレーヤの三人人狼における戦略を分析した [大澤 16]。

この研究を踏まえ、奈良女子大学の玉井日菜子は「3者間人狼における他者の投票行動を考慮した戦略の検討」 [玉井 19] において、大澤らのモデルを他者の投票行動を予想して自らの投票先を決定するように拡張し、分析を行った。具体的には、相手の投票行動を仮定したうえで、どのプレーヤに投票することが有利であるかを考察し、戦略を決定する。さらに、決定された戦略を踏まえ、こちらがその戦略をとるとすると予想した相手プレーヤが新たに戦略を決定する……ということを繰り返し行った場合の結果について検討を行っている。分析の結果、大澤らの研究では合理的とされた戦略に収束せず、一定間隔で同じ戦略が現れ、安定した戦略に収束しなかったことを報告した。

東京大学の武田惇史らは「五人人狼における戦略進化のシミュレーション」 [武田 19] において、三人人狼よりも複雑な五人人狼を対象とし、過去の人狼知能大会のログの分析、およびシミュレーションを行うことで戦略の進化を分析した。まず、2017年と2018年の人狼知能大会のログを用い、エージェントの採用した戦略がどのように遷移するのかを回帰分析を用いて確認した。その結果、2018年大会のエージェントは2017年大会のエージェントに対して勝率が高くなるような戦略を採用していることを明らかにした。次に、五人人狼のゲームを現実的な時間で網羅的かつ決定論的にシミュレーション可能のようにモデル化し、戦略の遷移を分析した。ただし、モデル化に際し、2日目に起こることはいっさい考慮せず、1日目の時点で決着がつくものとするなど、実際のルールとは異なるいくつかの制約を設けている。シミュレーションの結果、戦略の遷移を繰り返しても特定の戦略には収束せず、周期的に変化し続けるようになるという結果を得た。また、本シミュレーションは2017年と2018年のエージェントが採用した戦略の進化を内包していることが確認された。

3.4 人狼ゲーム以外の研究

本OSでは、人狼ゲーム以外の研究発表も2件行われた。

武蔵野大学の渡邊紀文らによる「協調パターンタスクの行動実験に基づいたシミュレーション分析」 [渡邊 19] では、二次元グリッド上で4名がそれぞれ一つずつコマを操作し、3コマからなる目標パターンを言語コミュニケーションを行わずに達成するタスクを対象とし、被験者実験によるデータ収集とデータ分析をもとにモデル化したエージェントによるシミュレーション実験を行った。本研究で扱った協調パターンタスクでは、プレーヤ間の言語コミュニケーションは存在しないものの、他のプレーヤがどう行動するかを予想したうえで自分の行動を決定するという他者の意図のモデル化が必要となるという点で、人狼ゲームと共通している。本研究においてもその戦略を有するエージェントモデルを作成し、シミュレーションを行ったところ、被験者実験と同様の結果が得られたと報告している。

筑波大学の佐藤栄介らは「相手の思考時間の長短によって推定の信頼度を変更する協力ゲーム Hanabi のエージェントの開発および評価」 [佐藤 19] と題し、Hanabiをプレイするエージェントを構築し、被験者実験を行った結果について報告した。Hanabiは協力型のカードゲームであり、本研究では二人でプレイする場合を対象としている。Hanabiの大きな特徴は、自分の手札は見るができず、相手の手札のみを見ることができると、相手への情報提供の内容がトークンによって制限されている点である。本研究では、相手の思考時間から相手の行動選択の自身の有無を推定し、それに基づいて自分の行動を決定するエージェントを構築している。被験者実験の結果、従来の戦略を用いたエージェントと比べ、得点と推定の成功率に有意差は見られなかったが、推定の成功数は増加したことを報告している。

4. 人狼知能をこれから始めるには

本稿により人狼知能に興味をもたれた方は、人狼知能を始めるためのさまざまな資料が公開されているのでぜひ参考にされたい。

まず、人狼知能プロジェクトについて説明した解説記事「人狼知能プロジェクト」が本誌に掲載された [片上 15]。人狼とは何かについての基本的な話から、人狼の基本戦略、人狼知能プロジェクトとAI研究の重要性、そして将来展望の話まで、非常にわかりやすく解説している。その後、人狼知能研究を行うための解説として、「人狼知能研究のすすめ」が「知能と情報」に掲載された [片上 18]。こちらは、最新の人狼知能の研究に関する情報と、人狼知能大会の現状、これまでのプロトコル部門に加えて、自然言語で人狼を行うAIの大会である自然言語部門とその評価方法などについて詳しく説

明した解説になっている。

まとまった書籍としては、森北出版「人狼知能～だます・見破る・説得する人工知能～」[鳥海 16]を読んでほしい。人狼知能とは何か？ から始まり、人狼ゲームの解説と人狼をプレイする AI システム研究の概要や超えるべき課題、そして人狼知能による未来まで、幅広く人狼知能について学べる一冊になっている。「だます」、「見破る」、「説得する」といった、人間ならではのコミュニケーションが戦略の鍵を握る人狼ゲームを人工知能に行わせようとする我々のわくわく感が伝わるのでは、と思っている。

自分で人狼知能を作成してみたい方は、マイナビ出版「人狼知能で学ぶ AI プログラミング」[狩野 17]が最適だと思われる。Java を使った人狼知能の作成の仕方から、機械学習として SVM (サポートベクタマシン) を活用した人狼知能の開発の仕方、自然言語処理を用いた人狼知能の開発の仕方など、人狼ゲームを用いて具体的かつ実践的なプログラミングの方法を教えており、楽しみながらプログラミングを学びたい人にうってつけの教科書になっている。

最新の情報、イベント、プログラムなどは、人狼知能プロジェクトの HP (<http://aiwolf.org/>) を確認してほしい。開発に関する情報は、例えば、開発関連のページに、以下のようなスライドや、ソースコードが誰でもダウンロードできるようになっている。

- 人狼知能エージェントの作り方 (Java)
- C# 版人狼の作り方
- AIWolfPy (Python 用プラットフォーム)

また、人狼知能大会では、決勝進出チームには、ソースコードの提出をお願いしている。過去の人狼知能大会の決勝進出者のチームソースコードが置いてあるので、現在の最強人狼知能のソースを参考に、さらに強力な独自の人狼知能を開発してほしい。これにより、人狼知能コミュニティによって、だます、見破る、説得するといった高度なコミュニケーションを可能とする汎用的な人工知能の研究の発展に寄与できることを期待する。

5. 第 1 回人狼知能国際大会の報告

冒頭で記載したとおり、今年初めて第 1 回の人狼知能国際大会が IJCAI 2019@マカオにて開催されたので、ここで簡単に紹介する。会場の様子は図 2 に示す。今回の大会では 94 チームもの登録があり、74 体のエージェントが投稿され、31 体が予選で稼働して対戦を行うことができた。最終的に 12 チームによって決勝が行われた。各エージェントは、最低 10 000 回以上の対戦を行い、その勝率で第 1 回国際大会の順位を決定した。すでに日本国内で 4 回の大会が行われており、継続して参加している人狼知能のレベルも年々上がっている。セッション内では大会に出場した人狼知能において使われた



図 2 第 1 回人狼知能国際大会 (AI Wolf Competition in IJCAI 2019)



図 3 第 1 回人狼知能国際大会表彰式の様子

手法についても紹介が行われた。本大会では機械学習を用いたチームが多く決勝に進出し、ベイズ推論、強化学習、決定木、LSTM、CFR (Counterfactual Regret minimization) などさまざまな手法が利用されていた。最後に、入賞した参加者にはその場で表彰が行われた。授賞式の様子を図 3 に示す。

6. おわりに

本稿では、2019 年度人工知能学会全国大会 (第 33 回) のオーガナイズドセッション「人狼知能と不完全情報ゲーム」における研究発表の紹介を通じ、現在の人狼知能研究の動向の解説を行った。また、IJCAI 2019 で開催された第 1 回人狼知能国際大会についても合わせて紹介した。

人狼知能に関する研究はゲーム情報学やインタラクション、自然言語処理などさまざまな研究会・学会などで散発的に発表が行われており、人狼知能に関わる研究者が一堂に会し、幅広い視点から議論を行うことができた本 OS は非常に有意義であったと感じている。来年のオーガナイズドセッションに採択されれば 3 回目でも最後となる予定であるが、より多くの方に参加いただき、有意義な議論の場となることを期待したい。

◇ 参考文献 ◇

- [Braverman 08] Braverman, M., Etesami, O. and Mossel, E.: Mafia: A theoretical study of players and coalitions in a partial information environment, *The Annals of Applied Probability*, pp. 825-846 (2008)
- [Chittaranjan 10] Chittaranjan, G. and Hung, H.: Are you awerewolf? Detecting deceptive roles and outcomes in a conversational role-playing game, *IEEE Int. Conf. on Acoustics Speech and Signal Processing*, pp. 5334-5337 (2010)
- [福田 19] 福田宗理, 穴田 一: 人狼ゲームにおける会話情報による役職推定, 2019年度人工知能学会全国大会(第33回) 論文集(2019)
- [狩野 17] 狩野芳伸, 大槻恭士, 園田亜斗夢, 中田洋平, 箕輪 峻, 鳥海不二夫: 人狼知能で学ぶ AI プログラミング: 欺瞞・推理・会話で不完全情報ゲームを戦う人工知能の作り方, マイナビ出版(2017)
- [片上 15] 片上大輔, 鳥海不二夫, 大澤博隆, 稲葉通将, 篠田孝祐, 松原 仁 ほか: 人狼知能プロジェクト(特集: エンターテインメントにおける AI), *人工知能*, Vol. 30, No. 1, pp. 65-73 (2015)
- [片上 18] 片上大輔, 鳥海不二夫, 大澤博隆, 狩野芳伸, 稲葉通将, 大槻恭士: 人狼知能研究のすすめ(特集: 人間共生システムおよびその応用), *知能と情報*, Vol. 30, No. 5, pp. 236-244 (2018)
- [小村 19] 小村友希, 坂本 航, 尾崎知伸: 人狼ゲームにおける明示的役職・陣営推定理由の抽出, 2019年度人工知能学会全国大会(第33回) 論文集(2019)
- [Migdal 10] Migdal, P.: A mathematical model of the Mafia game, *Arxiv preprint arXiv:1009.1031* (2010)
- [大川 17] 大川貴聖, 吉仲 亮, 篠原 歩: 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* (2017)
- [大澤 16] 大澤博隆, 佐藤 健: 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論文集(第30回) (2016)
- [王天 18] 王天 鶴, 金子知適 ほか: 人狼エージェントにおける深層 Q ネットワークの応用, *ゲームプログラミングワークショップ 2018 論文集*, Vol. 2018, pp. 16-22 (2018)
- [佐藤 19] 佐藤栄介, 大澤博隆: 相手の思考時間の長短によって推定の信頼度を変更する協力ゲーム Hanabi のエージェントの開発及び評価, 2019年度人工知能学会全国大会(第33回) 論文集(2019)
- [高山 19] 高山周太郎, 大澤博隆: 実世界人狼ゲームのジェスチャーの分析, 2019年度人工知能学会全国大会(第33回) 論文集(2019)
- [武田 19] 武田惇史, 鳥海不二夫: 5 人狼における戦略進化のシミュレーション, 2019年度人工知能学会全国大会(第33回) 論文集(2019)
- [玉井 19] 玉井日菜子: 3者間人狼における他者の投票行動を考慮した戦略の検討, 2019年度人工知能学会全国大会(第33回) 論文集(2019)
- [鳥海 16] 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 (2016)
- [渡邊 19] 渡邊紀文, 糸田孝太: 協調パターンタスクの行動実験に基づいたシミュレーション分析, 2019年度人工知能学会全国大会(第33回) 論文集(2019)

- [Xia 07] Xia, F., Wang, H. and Huang, J.: Deception detection via blob motion pattern analysis, *Affective Computing and Intelligent Interaction*, pp. 727-728 (2007)
- [山本 19] 山本浩隆, 棟方 渚: 生体信号に基づく人狼プレイヤーの「強さ」分析の試み, 2019年度人工知能学会全国大会(第33回) 論文集(2019)
- [Yao 08] Yao, E.: A theoretical study of mafia games, *Arxiv preprint arXiv:0804.0071* (2008)
- [Zhou 08] Zhou, L. and Sung, Y.: Cues to deception in online Chinese groups, *Proc. 41st Annual Hawaii Int. Conf. on System Sciences*, pp. 146-146 (2008)

—— 著者紹介 ——



稲葉 通将 (正会員)

2012年名古屋大学大学院情報科学研究科博士後期課程修了。同年、広島市立大学大学院情報科学研究科助教, 2019年電気通信大学人工知能先端研究センター准教授, 現在に至る。博士(情報科学)。対話システム, 対話処理に関する研究に従事。電子情報通信学会, 情報処理学会, 言語処理学会各会員。



片上 大輔 (正会員)

2002年東京工業大学大学院総合理工学研究科知能システム科学専攻博士後期課程修了。博士(工学)。同年、東京工業大学大学院総合理工学研究科助手。2007年同研究科助教。2006年英国ハートフォードシャー大学, スイス・チューリッヒ大学の客員研究員を兼任。2010年東京工芸大学工学部コンピュータ応用学科准教授, 2017年同教授, 現在に至る。日本知能情報ファジィ学会, IEEE, ACM 各会員。HAI に関する研究に従事し, 人狼知能および雰囲気工学に興味をもつ。本学会理事。



狩野 芳伸 (正会員)

2001年東京大学理学部物理学科卒業。2007年同大学院情報理工学系研究科博士課程単位取得退学。2011年同研究科にて博士(情報理工学)。同研究科特任研究員, 科学技術振興機構さきかけ研究者などを経て, 2014年より静岡大学情報学部准教授。言語処理学会, 情報処理学会各会員。



大槻 恭士 (正会員)

1993年東北大学大学院工学研究科情報工学専攻博士後期課程修了。博士(工学)。現在, 山形大学学術研究院准教授。統計的言語モデル, 音声情報処理, 知能ロボティクスなどの研究に従事。情報処理学会, 日本音響学会などの会員。