

平成 30 年 5 月 31 日現在

機関番号：25403

研究種目：挑戦的萌芽研究

研究期間：2015～2017

課題番号：15K12180

研究課題名(和文)人狼ゲーム研究：他者中の自己モデルを考えた説得的人工知能の開発に向けて

研究課題名(英文)A study of the werewolf game : Toward the development of persuasive artificial intelligence considering self model in other player

研究代表者

稲葉 通将 (Inaba, Michimasa)

広島市立大学・情報科学研究科・助教

研究者番号：10636202

交付決定額(研究期間全体)：(直接経費) 2,700,000円

研究成果の概要(和文)：本研究の最終目的は、人と一緒に人狼ゲームをプレイできるエージェントの実現にある。研究期間内にはその第一歩として、人狼ゲーム専用のプロトコルを設計し、それをを用いてコミュニケーションを行うエージェントに関する研究を実施した。特に、自分から他者がどう見えるかという「自己の心理モデル」に加え、他者から見て自分がどう見えるかという「他者から見た自己のモデル化」の2つの階層的な心理モデルを持つエージェントを構築した。また、ゲーム研究においては、多くの人が参加して競い合うことで、集合知を集められる環境の整備が重要である。そこで研究開発基盤を整備し、一般参加可能なコンピュータ人狼大会を開催した。

研究成果の概要(英文)：Our final objective is making a werewolf player agent that can play with human players. In this study, as a first step, we designed a protocol for enabling agents to have a conversation and conducted studies on agents that communicate using the protocol. We propose a psychological model that considers multiple perspectives to model the play of a human such as inferring the intention of the other side. As one of the psychological models, we constructed "self model" that models the role of others as viewed from their own viewpoint and "others model" that models the role of others as viewed from their viewpoints. Combining these models, we constructed an agent with hierarchical psychological models.

In the study of game informatics, it is important to develop an environment that collects collective intelligence. We developed a R&D infrastructure and held computer werewolf game competitions.

研究分野：対話型ゲーム

キーワード：人狼ゲーム 人狼知能 対話システム

1. 研究開始当初の背景

ゲーム情報学の技術発展は、社会的なインタラクションだけでなく、並列計算や情報探索などの技術にも多大な貢献をもたらしてきた。近年、コンピュータはチェス、将棋、囲碁などの二人完全情報ゲームにおいて人間を圧倒しつつあり、次の研究対象となるゲームが模索されている。

そこで本研究ではコミュニケーションにより勝敗が決定するゲーム「人狼ゲーム」に着目した。人狼ゲームの持つ一般性・普遍性の高さから、本研究が貢献できる領域はボードゲーム研究が主に対象としてきた探索手法や機械学習に留まらない。エージェントインタラクション、言語処理、認知科学や心理学など含む領域が、人狼知能研究の対象とする領域である。人狼ゲームでは論理的な思考により相手の意図を読むことはもちろん、説得を行い、信頼を獲得することが極めて重要である。これらは我々が社会活動において日常的に行っていることでもあり、様々な対話エージェントに適用可能な汎用コミュニケーション技術である。このような高度なコミュニケーション技術がゲームの勝利のためには必須であり、かつ勝敗による客観的な評価が可能である人狼ゲーム研究は、技術の発展を大きく後押しできる可能性がある。

2. 研究の目的

本研究の最終目的は、身体性を持ち、音声会話によって人狼ゲームをプレイできるエージェントの実現にある。研究期間内にはその第一歩として、処理の難しい自然言語ではなく、人狼ゲーム専用のプロトコルを設計し、それを用いてコミュニケーションを行うエージェントに関する研究を実施した。また、ゲーム研究においては、様々なツールやデータを積極的に公開することで参入障壁を下げ、多くの人々が参加して競い合うことで、集合知を集められる環境の整備が重要である。そこで注釈付きゲームログの構築を含む研究開発基盤を整備し、一般参加可能なコンピュータ人狼大会を開催した。

3. 研究の方法

本研究で設定した目的に対し、「人狼ゲームのプレイログの分析」、「コンピュータ人狼大会の実施およびそのための環境整備」、「階層的な心理モデルを用いたエージェントの開発」の3点を実施した。

(1) 人狼ゲームのプレイログの分析

本研究では、Web上でオンラインゲームとして公開されている人狼BBSのログを使用し、勝利のために有効なコミュニケーション戦略を明らかにすることを目指す。そのような戦略が明らかとなれば、強い人狼知能を構築する際の指針とすることができる。また、コミュニケーションとプレイヤーの行動の関係を分析することも本研究の目的である。

人狼ゲームでは、プレイヤーの主観(他のプレイヤーに対する印象など)も行動に大きく影響する。そこで、プレイヤーの主観に大きく影響を与えるコミュニケーションと行動の関係を分析することで、客観的な情報だけでなく、主観的な情報も考慮できる人狼知能の構築に役立つ知見を得ることができる。これは、ただ強いだけでなく、人間らしい思考の実現や、主観的な情報を重視する度合いを変えて個性を出すなど、人が一緒にプレイして楽しい人狼知能の実現のために有用な知見である。

そこで、本論文では、ゲームの勝利のために有効なコミュニケーション戦略を明らかにするとともに、プレイヤーの行動の関係を分析するため、発話内容を表すタグの設計とゲーム内の発話へのタグ付与とタグを用いた分析を行った。

(2) コンピュータ人狼大会の実施およびそのための環境整備

人狼ゲーム研究における開発基盤構築、およびコンピュータ人狼大会実施のため、人狼ゲームエージェント開発・対戦用フレームワークの開発を行う。なお、本フレームワークではこれまでに提案した人工知能により人狼をプレイするためのコミュニケーション言語(プロトコル)を用いたエージェントを対象とする。

また、人間同士のゲームログの分析を容易にするため、人狼BBSにおけるゲームに対し、アノテーションを行い、コーパス構築を行う。

(3) 階層的な心理モデルを用いたエージェントの開発

人狼ゲームでは会話以外の客観的な情報がほとんど存在しないため、相手の立場に立って考え、相手の意図を推理する必要がある。そこで本研究では、自分から他者がどう見えるかという「自己の心理モデル」に加え、他者から見て自分がどう見えるかという「他者から見た自己のモデル化」の2つの階層的な心理モデルを持つエージェントの構築を行う。また、最適化した心理モデルを用いて他者が誰に投票しようとしているかを予測し、今日自分が誰に投票すれば他者からの投票を避けることができるかを考えて投票先を決定するエージェントの構築を行い、対戦実験により評価する。

4. 研究成果

(1) 人狼ゲームのプレイログの分析

本研究では、人狼BBSのゲームデータを分析し、人狼ゲームにおけるプレイヤーの発話内容を表現する20種類のタグを設計した。次に、設計したタグを人狼BBSにおけるプレイヤーの発話に付与し分析を行った。タグ付与を行ったゲーム数は24個であり、総発話数は19,263個である。分析では、襲撃対象、および追放対象の決定にプレイヤーごとの

発話の傾向がどのように影響するのかを調査した。また、ゲーム全体の発話の傾向とゲームの勝敗の関係について分析した。分析の結果、人間側、人狼側の各プレイヤーが自陣営の勝利のために効果的なコミュニケーション戦略、および特定のコミュニケーションとプレイヤーの行動の関係が明らかとなった。例えば人間側のプレイヤーは「占いに關することを最優先で話し合う」ことが勝利のために重要であり、人狼側は「雑談をしない」ことが重要であることなどが確認された。

(2) コンピュータ人狼大会の実施およびそのための環境整備

A. 人狼知能プラットフォームの開発

人狼知能プロトコルを用いたコミュニケーションが可能な人狼ゲームエージェントが作成・対戦可能なプラットフォームの構築を行った。開発言語は Java とし、C#での開発についても山形大の大槻准教授の協力を得ることができた。本プラットフォームは「人狼知能プラットフォーム」という名称で人狼知能プロジェクトの Web ページで公開している (<http://aiwolf.org/server>)。また、本プラットフォームを用いて作成したエージェントと人間が対戦することが可能なソフトも開発し、同じく公開している。

B. コンピュータ人狼大会の実施

2015 年の 8 月には、本プラットフォームを用いて作成したエージェントのコンテストである「第一回人狼知能大会」を開催した。2016 年、2017 年にも同じく大会を実施し、毎回 30 チーム以上の参加がある。2018 年にも 8 月に開催が決定している。なお、これまでの決勝進出チームのソースコードはすべて人狼知能プロジェクトのホームページで公開中である。

また、2017 年の大会からは、これまでと同じく人狼知能プロトコルを用いた「プロトコル部門」に加え、自然言語（日本語）により対戦を行う「自然言語部門」を新たに設定し、大会を実施した。自然言語部門ではエージェントの強さではなく、コミュニケーションの自然さなどを複数の判定者により評価した。2018 年の大会でもプロトコル部門と自然言語部門の 2 部門制で大会を実施する。

C. アノテーション付きコーパスの公開

人狼知能大会において自然言語部門が開始されたが、自然言語により自然な議論が可能なエージェントの実現には、発話意味理解、発話生成など多くの課題が存在する。しかし、その学習のためのコーパスなどは整備されておらず、統計的な手法を適用することは困難である。そこで、人狼 BBS のログに対しアノテーションを行い、コーパスの構築を行った。アノテーションとして、役職表明と能力行使報告情報の 2 点をゲーム中のプレイヤーの発話に付与した。アノテーションはクラウ

ドソーシングにより行い、アノテーション作業とそのチェックの作業の 2 段階により、誤りの少ないアノテーションを目指した。現在までに 300 ゲームのアノテーションを完了しており、コーパスは Web 上で公開している。

(3) 階層的な心理モデルを用いたエージェントの開発

自分が他のプレイヤーをどの役職と考えているかという「自己モデル」と他のプレイヤーが他のプレイヤーをどの役職と考えているかという「他者モデル」の両方を階層的に用いて適切な行動を決定するエージェントを開発した。自己モデルと他者モデルにおけるパラメータは人狼経験者に対し、人狼ゲームのある場面における行動のアンケートを取ることで決定した。実験の結果、自己モデルのみのエージェントと比較し、自己モデルと他者モデルの両方を持つエージェントが有意に性能が高いことを確認した。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 3 件)

Fujio Toriumi, Hiroataka Osawa, Michimasa Inaba, Daisuke Katagami, Kosuke Shinoda, and Hitoshi Matsubara: AI Wolf Contest -Development of Game AI Using Collective Intelligence-. Communications in Computer and Information Science, Computer Games. Springer, Cham, pp.101-115, 2017.

稲葉通将, 大畠 菜央実, 鳥海不二夫, 高橋健一: 雑談ばかりしていると殺される? 人狼ゲームにおける発話行為タグセットの提案とプレイヤーの行動・勝敗の分析. 情報処理学会論文誌, Vol.57, No.11, pp.2392-2402, 2016.

〔学会発表〕(計 14 件)

Hirofumi Nakamura, Daisuke Katagami, Fujio Toriumi, Hiroataka Osawa, Michimasa Inaba, Kosuke Shinoda and Yoshinobu Kano: Generating human-like discussion by paraphrasing a translation by the AIWolf Protocol using Werewolf BBS Logs. IEEE conference on Fuzzy Systems 2017(FUZZ-IEEE2017), 2017.

狩野芳伸, 稲葉通将: 人狼知能大会における自然言語処理部門の構想. 言語処理学会第 23 回年次大会, 2017.

Noritsugu Nakamura, Michimasa Inaba, Kenichi Takahashi, Fujio Toriumi,

Hiroataka Osawa, Daisuke Katagami, Kosuke Shinoda: Constructing a Human-like agent for the Werewolf Game using a psychological model based multiple perspectives. 2016 IEEE Symposium Series on Computational Intelligence (SSCI2016), 2016.

Yuya Hirata, Michimasa Inaba, Kenichi Takahashi, Fujio Toriumi, Hiroataka Osawa, Daisuke Katagami, Kosuke Shinoda: Werewolf Game Modeling using Action Probabilities based on Play Log Analysis. The 9th conference on Computer and Games (CG2016), 2016.

栢野航, 工藤佑介, 大澤博隆, 片上大輔, 鳥海不二夫, 稲葉通将, 篠田孝祐: 人狼を演じるロボットエージェント観戦による社会的知能の学習. HAI シンポジウム 2015, 2015.

中村洋文, 片上大輔, 鳥海不二夫, 大澤博隆, 稲葉通将, 篠田孝祐: 人狼 BBS のログデータを用いた AI の自然な議論空間の構築. Joint Agent Workshops and Symposium 2015 (JAWS2015), 2015.

〔図書〕(計 2 件)

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 森北出版, 2016

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

〔その他〕

ホームページ等

人狼知能プロジェクト <http://aiwolf.org/>

Aiwolf (GitHub) 開発したツール・コーパス
<https://github.com/aiwolf>

6. 研究組織

(1) 研究代表者

稲葉 通将 (INABA, Michimasa)
広島市立大学・情報科学研究科・助教
研究者番号: 10636202

(2) 研究分担者

大澤 博隆 (OSAWA, Hiroataka)
筑波大学・システム情報系・助教
研究者番号: 10589641

鳥海 不二夫 (TORIUMI, Fujio)
東京大学・大学院工学系研究科・准教授
研究者番号: 30377775

片上 大輔 (KATAGAMI, Daisuke)
東京工芸大学・工学部・准教授
研究者番号: 90345372

篠田 孝祐 (SHINODA, Kosuke)
電気通信大学・大学院情報システム学研究科・助教
研究者番号: 90533191