

# 信頼度によって説得行動を変化させる人狼知能

野中 和馬<sup>1,a)</sup> 松原 仁<sup>2</sup>

**概要:** 人工知能の研究題材として人狼ゲームが注目されている。先行研究において役職推定は多いが説得を重視した研究はあまり多くない。そこで説得に重点を置いたゲーム AI 作成を目的とする。そのために信頼度というものを実装、利用することで説得行動を変化させることを目標とした。現状では信頼度の実装を完了し説得等に利用する部分を制作している。

**キーワード:** ゲーム AI, 信用, 行動変化

## Werewolf Intelligence that Changes Persuasion Behavior Depending on Reliance Level

Nonaka Kazuma<sup>1,a)</sup> Matsubara Hitoshi<sup>2</sup>

**Abstract:** Werewolf games have been attracting attention as a subject of research in artificial intelligence. In the previous studies, there are many studies on estimating position, but not many studies on persuasion. The goal of this research is to change persuasion behavior by implementing and using the reliance level. Currently, we have completed the implementation of the reliance level and are working on the part to be used for persuasion.

**Keywords:** Game AI, Trust, Behavior Change

### 1. はじめに

AI にゲームをさせる試みは多くされている。例えば将棋では Ponanza, 囲碁では alphaGo など人間を超えるゲーム AI が生まれている。しかし、まだまだ発展の見込みのあるゲーム AI はある。その中でも人狼ゲームは対話を重視しており、人工知能の対人技術への応用ができる可能性があり注目されている。人狼ゲームを行うゲーム AI (人狼知能) の開発は人狼知能プロジェクト[1]を主体として多く行われ、研究もされている。また、人狼知能作成を用いたプログラミングの本も出版されている。[2]しかし、説得を重視した研究はあまり多くはない。そこで本研究では説得することに重点を置いた人狼知能の開発を目的とした。具体的には信用度を導入し説得行動に変化をつけ味方を増やす方法を模索する。

### 2. 人狼ゲーム

#### 2.1 人狼ゲームの概要

ここでは人狼ゲームのルールについて説明をする。人狼ゲームでは「村人陣営」と「人狼陣営」の2陣営に分かれ、それぞれ割り当てられた役職の能力を使い情報を得る。その情報をもとに話し合いを行い参加者の中に紛れている人狼を見つけ出し、村から追放することを目的とするゲームである。

#### 2.2 人狼ゲームの勝利条件

村人陣営は村から人狼と呼ばれる役職の人を追放できれば勝利である。一方、人狼陣営は生存者の半数以上が人狼となれば勝利である。

#### 2.3 人狼ゲームの流れ

このゲームは3フェイズに分かれており、それぞれ「昼時間」「投票時間」「夜時間」と呼ぶこととする。昼時間には参加者たちの中で誰が人狼なのか議論を行う時間である。投票時間は昼時間の議論を踏まえて、各参加者が怪しいと感じた参加者に投票を行い、得票数の最も多かった参加者をゲームから追放する。夜時間には与えられた役職の人々が能力を使用する

1 公立はこだて未来大学院 システム情報科学研究科  
Graduate School of Future University Hakodate, The University  
of Game System Informatics

2 公立はこだて未来大学 Future University Hakodate  
a) g2120039@fun.ac.jp

ことができる.このフェイズを繰り返し,どちらかの陣営が勝利するまで続ける.

役職と能力について表1に示す.

表1 役職と能力

役職名	能力	陣営
村人	特に能力はない。	村人
占い師	参加者1人を指定してその人が人狼かそうでないかが分かる。	村人
霊媒師	その日投票で追放された人が人狼かそうでないかが分かる。	村人
人狼	参加者1人を指定してその人を追放する。 人狼同士は秘密裏に会話できる。	人狼
狂人	特に能力はない。 人狼陣営の村人。	人狼

### 3. 関連研究

人狼知能には相手の発言から役職を推定する研究がある[3].この研究では従来のモデルを参考に発言から発言相手や対象となった参加者の役職を推定する機能を追加している.結果として従来モデルよりも一定の状況では精度が向上している.発言から役職を推定する機能は本研究における信頼度の見極めにも応用ができると考えている.

また,説得に関する人狼知能の研究例を挙げる.大抵らは既存の人狼知能に占い結果によって説得を行わせるプログラムを作成している[4].その結果として勝率の向上が認められた.この結果から説得機能は人狼知能の機能向上に影響があると考えられる.

## 4. 説得行動について

### 4.1 人狼ゲームにおける説得行動の必要性

人狼ゲームにおいてなぜ説得行動が必要なのかを説明する.人狼ゲームでは話し合いを行い,その後追放する人を決定する手段として多数決を用いる.生き残っている人数や内訳で勝負が決まる人狼ゲームでは誰が追放されるかが大きな要因となる.つまり,話し合いの場で多くの味方を付け,自身に有利な意見を通すかが重要な点であり,そのためには説得を行う必要が出てくる.また,人狼知能に関する書籍でも説得行動について触れられており,重要な点であると考えられ

る.[5]

### 4.2 本研究における説得行動の定義

本研究における説得行動とは何かを説明する.説得とは相手と話し,納得させることである.しかし,一言に説得といっても対象や目的は様々である.対立する相手の意見を改めさせて味方にすることも説得であり,同じ意見の人を敵にならないように固定することも説得といえる.

本研究での説得行動とは,同じまたは似た意見の人を味方にするものと定義する.理由として,人狼ゲームにおいて時間や発言回数といった制限がある中で意見の対立する相手を味方に変えることは困難であるからである.対して,既に同じような意見の人を固定し味方にするほうが比較的容易であり,費用対効果も高くなると考える.

### 4.3 人狼ゲームの説得行動例

人狼ゲーム内で行われる説得行動の例を挙げる.

ゲームの終盤に占い師,人狼,村人各1人の3人となった場合,占い師と人狼が村人に対して説得を行う状況が発生する.(図1)

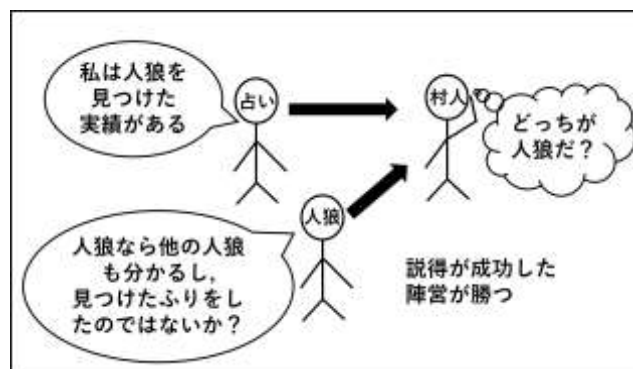


図1 説得行動の状況例

この状況では村人の説得に成功したほうが勝利することができる.説得方法として,これまでの功績を挙げ,自分のほうが貢献していたと主張する方法や,相手の言動に対して考察し,相手が信用ならないと主張する方法などが行われる.

### 5. 信頼度

ここでは実装している信頼度について説明をする.信頼度とは相手に対して自分がどれだけ信頼できると思っているかの度合いを表している.この信頼度の変化は相手の言動によって引き起こされる.例えば,自分に対して肯定的な行動をすれば信頼度は上がる.逆

に否定的,または敵対的な行動をとれば下がる. 具体的な増減例は表 2 にまとめる.信頼度には段階を作っており,中間を 5 として 0~10 まででの 11 段階で表現している.0 に近いほど信頼できないと判断している状態としている.逆に 10 に近いほど信頼できると判断している状態にある.初期状態は 5 に設定する.

表 2 信頼度の増減例

状況	増減
自分に疑いを向ける発言をする	-1
自分に投票を行う	-2
ルールの矛盾が生じる発言をする (自分が占いの場合に自分以外が自分は占いだと発言する等)	-10
自分の発言に対して肯定する	+1
説得した場合にその通りに動く	+2

## 6. 実装方法

ここでは実装方法について説明をする.本研究では一から人狼知能を作成するわけではなく,既存の人狼知能に前述した信頼度機能の追加,実装する.作成をおこなうために参考にした人狼知能は,人狼知能プロジェクトで開催されている全国大会で使用されたものである.信頼度については前述した通り言動を入力として信頼度という形で出力を行う.その信頼度を今度は入力として自身の説得行動に変化を持たせる.具体的には信頼度の高い相手には説得を優先的に行う.また,信頼度の高い参加者の意見に対して同調を示し,意見を強化するといった行動変化を考えている.逆に信頼度が低い参加者には説得行動を行わないようにする.そして,優先的に信頼度の低い相手が投票されるように周囲へ説得を行うようにする. (図 2)

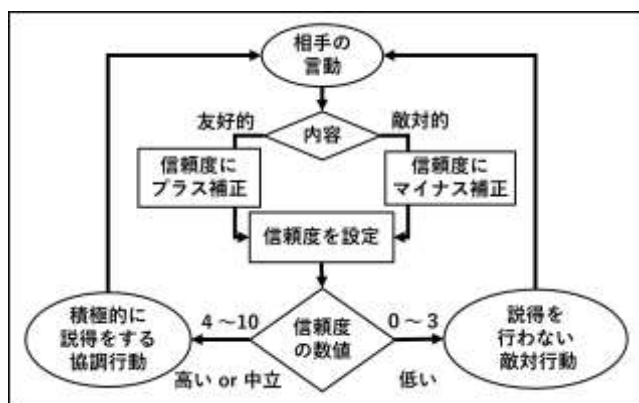


図 2 信頼度の変動と行動変化

この行動によって他参加者への言動に影響を与え,さらに信頼度に変化を生み出す.そうすることで敵と

味方をはっきり区別し,味方に対してはさらに強い信頼関係を構築し,敵を減らすことを期待する.

## 7. 評価方法

ここでは評価方法について説明をする.評価方法として,元となった人狼知能と説得機能を追加した人狼知能を使用して人狼ゲームを行う.人狼ゲームの設定は,参加者は 5 人,そして毎回参加者の役職を固定する.参加者の組み合わせによって A 組と B 組の 2 つに分ける.まず, A 組として 1 人は説得を行う人狼知能を使用し,その他の 4 人を元の人狼知能を使用する.そして, B 組として全員元の人狼知能を使用する. A 組と B 組それぞれでの行動を比較し考察をおこなおうと考えている.

参加者の役職と参加人数を固定する理由は,比較をやすくするためである.少人数であり,役職が固定であれば同じまたは似た状況を再現しやすいと考えた.

また,人数を増やすことで信頼度の与える影響が変化するか検証していきたい.

## 8. 今後の展望

現状では信頼度の設定が完了しているがこれを用いて説得行動に変化をもたらす機能はまだ構築中である.今後は説得行動の変化を完成させ実験,考察をする.そしてその結果から有効性があるのか検証していく.

## 参考文献

- [1] 人狼知能プロジェクトホームページ. <http://aiwolf.org/>.
- [2] 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, マイナビ出版, 2017.
- [3] 池尚子, 三品晟瑠, 山根健: 人狼ゲームにおけるプレイヤーの推測発言からの発言者および対象者の役職推定, 人工知能学会 第 34 回全国大会 (2020), DOI : [https://doi.org/10.11517/pjsai.JSAI2020.0\\_2F5OS20b03](https://doi.org/10.11517/pjsai.JSAI2020.0_2F5OS20b03).
- [4] 大槻恭士, 今田佑生也: 人狼知能エージェントにおける説得・被説得機能の評価, 人工知能学会 第 34 回全国大会 (2020), DOI : [https://doi.org/10.11517/pjsai.JSAI2020.0\\_1P3GS705](https://doi.org/10.11517/pjsai.JSAI2020.0_1P3GS705).
- [5] 鳥海不二夫, 片上大輔, 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.