

# 進化シミュレーションを用いた 3 人人狼の戦略分析

汪 博豪<sup>1</sup> 大澤 博隆<sup>1</sup> 佐藤 健<sup>2</sup>

<sup>1</sup>筑波大学 <sup>2</sup>国立情報学研究所

本研究では、人狼ゲームの最もシンプルな構成である 3 人人狼を考え、進化シミュレーションの手法を用いて、どのような現象が起きるか検討した。我々が定義した合理性に基づいてシミュレーションに用いた戦略を分類した。進化中の各エージェントの戦略がどのように変化していくか分析した。シミュレーションの結果により、不合理だと考えられた戦略が、ある時に最適な戦略である現象が起きた。プレイヤーが嘘をつくことにより、普段の不合理だと思われる戦略の選択率が変化することを確かめた

## 1 はじめに

不完全情報ゲームであり、正体隠匿型ゲームに分類されるコミュニケーションゲーム人狼を元にして、人工知能やエージェントなどの問題の解決のため、様々な研究が行われている[1][2][3]。人狼ゲームは、プレイヤー同士の情報が隠された騙し合いのゲームであり、多彩な戦略がプレイヤー間で行われるゲームである。

我々は人狼ゲームの戦略の中で、プレイヤーが意図的に事実とは異なる発言を行い、隠された味方のプレイヤーがその嘘を理解し不合理な戦略を選ぶ現象に注目した。人狼ゲームでは村人が人狼を出し抜くため、ゲームで使う用語にプレイヤー同士で通じる暗黙の意味が発生し、言語が移り変わっていくような現象がある。この現象は、コミュニティの中のサブコミュニティが意味を共有する言葉を作り出す、という点でいわゆる「若者言葉」の発生に似ている。「若者言葉」は、20 歳前後の青少年以外の世代ではあまり用いない言葉のことである。従って、村人陣営のみが分かる嘘は人狼ゲームの「若者言葉」だと考えられる。本研究では、上記の現象がどのように起きるかを検討するため、我々は以前の研究で提案した 3 人人狼の進化シミュレーションを行った[4]。シミュレーション結果を我々が定義した合理性に基づいて分類し、分析を行った。

## 2 3 人人狼の定義

### 2.1 人狼ゲームとは

人狼ゲームは嘘を見抜くコミュニケーションゲームの一種である。各プレイヤーは村人陣営と人狼陣営に分かれてプレイを行う。人狼ゲームにおける村人陣営には夜のフェーズに誰が人狼であるかを見分ける占い師という役職が存在する。人狼ゲームは昼

と夜のフェーズが交互に進み、以下の勝利条件を満たすことでゲーム終了する。昼フェーズでは自由議論を行い多数決投票による追放対象が決定される。夜フェーズでは人狼の襲撃、占い師の占いが行われる。すべての人狼が追放された場合に村人陣営が勝利となり、人狼の数が村人と同数以上になった場合に、人狼陣営の勝利になる。

### 2.2 3 人人狼ゲーム

上記人狼ゲームの最もシンプルな構成が、3 人人狼である。人狼ゲームはその性質上、2 人では遊べない。本研究では、村人および占い師は村人陣営となり、人狼が人狼陣営となる 3 人人狼ゲームを扱った。本研究では簡易化のため、人狼ゲームにおける昼フェーズでは、各プレイヤーが一回しか発言できない。人狼が村人を襲撃することで昼フェーズに進む前にゲーム終了してしまうので、本研究では人狼が襲撃できない。一方、占い師に事前にどちらが狼であるかの情報を与えてゲーム開始することを定義するため、夜のフェーズが存在しない。占い師は、誰が人狼であるかをゲーム開始前に知っている。そのため、いかなる状況であっても、占い師の追放先は必ず人狼である。従って、本研究で村人と人狼の投票先により、ゲーム結果は表 1 のとおりとなる。村人の選択が人狼投票になる場合に村人側ならず勝利する。占い師に投票を 2 票集められる場合に人狼側が勝利する。これ以外の場合、ゲーム結果は引き分けとなる。

表 1 : ゲーム結果の分類

Villager's Vote	Werewolf's Vote	Game Result
Werewolf	Seer	Villager wins
Werewolf	Villager	Villager wins
Seer	Seer	Werewolf wins
Seer	Villager	Draw

### 3 シミュレーションの説明

#### 3.1 会話の定義

人狼ゲームにおける最も難しい要素が、会話の定義である。人狼ゲームでは、自由会話を行われている。シミュレーションでは自由会話の完全対応が難しい。従って、本研究で扱う会話形式は、下記の3通りである。

1. 「私は村人である」
2. 「私は占い師であり、左のプレイヤーが人狼だ」
3. 「私は占い師であり、右のプレイヤーが人狼だ」

#### 3.2 シミュレーションの手続き

我々は以前の研究で提案した各プレイヤーが合理的な初期手[5]を設定された3人人狼の進化シミュレーションを行い、どのような戦略があるかどのような現象が起きるかを求めた。手続きを図1に示す。各プレイヤーはまず、自身の内部状態からはツン現内容を決定する(genecode1)。入力となる内部状態は2種類であり、発言は3種類である。次に、各プレイヤーは他のプレイヤーの2つの発言から、自身の内部状態を更新する(genecode2)。他プレイヤーの各発言は3通りであるため、genecode2の入力は9種類、更新する内部状態は2種類となる。

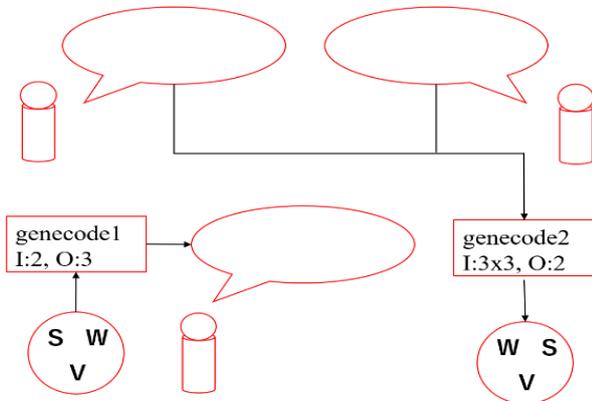


図1：シミュレーションのプロセス

進化されたエージェントの数により、3条件に分けて7種類の進化シミュレーションを行った。各種類において、エージェントを10000世代に渡って進化させた。各世代において、エージェントはランダムで5000試合を行う。試合後、勝率、嘘をついた割合、占い師としてCOした割合の計算を行った。各種類のシミュレーションを100回行い、その結果の平均値をとって分析を行った。

### 4 戦略の分類

本研究では、占い師が嘘をついているかを予想す

ることで村人と狼の戦略を、「合理的(rational)」や「不合理的(irrational)」や「あいまい(ambiguous)」に分類した。例えば、占い師としてCOしたプレイヤーの

表2：村人の戦略分類

P1's talk			P2's talk			D	P0's status
P1	P2	P0	P2	P0	P1		
V			V			0	Ambiguous
						1	Ambiguous
S	V	W	V			0	Irrational
						1	Rational
S	W	V	V			0	Rational
						1	Irrational
V			S	V	W	0	Irrational
						1	Rational
S	V	W	S	V	W	0	Irrational
						1	Rational
S	W	V	S	V	W	0	Ambiguous
						1	Ambiguous
V			S	W	V	0	Rational
						1	Irrational
S	V	W	S	W	V	0	Ambiguous
						1	Ambiguous
S	W	V	S	W	V	0	Rational
						1	Irrational

表3：人狼の戦略分類

P1's talk			P2's talk			D	P0's status
P1	P2	P0	P2	P0	P1		
V			V			0	Ambiguous
						1	Ambiguous
S	V	W	V			0	Irrational
						1	Rational
S	W	V	V			0	Ambiguous
						1	Ambiguous
V			S	V	W	0	Ambiguous
						1	Ambiguous
S	V	W	S	V	W	0	Irrational
						1	Rational
S	W	V	S	V	W	0	Ambiguous
						1	Ambiguous
V			S	W	V	0	Rational
						1	Irrational
S	V	W	S	W	V	0	Ambiguous
						1	Ambiguous
S	W	V	S	W	V	0	Rational
						1	Irrational

発言内容を考え、自分が村人として指されたらそのプレイヤーを占い師と考える場合、それは合理的な戦略である。これに対し、狼と考える場合に、それは不合理な戦略である。両方のプレイヤーの発言内容が同様である場合に、あいまいな戦略だと定義する。

## 5 シミュレーション結果による戦略分析

村人陣営のみ進化する場合に、村人エージェントが用いた戦略の割合を図2に示す。村人陣営勝利の時に用いた不合理な戦略が定期的に出現する。全エージェント進化する場合には、村人エージェントが用いた戦略の割合を図3に示す。村人の戦略が占い師の嘘の割合により、変化している。村人陣営勝利の時に用いた不合理な戦略が一時的に出現していることがわかる。

### 5.1 考察

シミュレーション結果より、予想していた占い師が

嘘をつき、村人が不合理な戦略を選んで占い師を見つける「若者言葉」に類似の現象が出現していた。村人陣営のみ進化する場合に、村人と占い師が互いに相手の思考モデルを学習でき、占い師が村人騙り、村人が村人騙りの占い師の正体を常に見つけて村人陣営の勝率を確保している。全エージェントが進化する場合に、狼が村人騙りを始めたため、狼の発言内容による村人が占い師の意図を常に把握することができなかった。従って、村人陣営の「若者言葉」に該当する現象が一時的に出現していた。

## 6 結論

本研究では3人狼ゲームのモデル化を行った後、「若者言葉」に類似したコミュニティ間のサブコミュニティでの意味の生成現象がどのように起きるか、進化シミュレーションの手法を用い、戦略を分類した分析を行った。分析結果より、村人陣営のみ進化する場合に村人陣営の「若者言葉」が安定に出現している。全エージェント進化する場合に、村人陣営の「若者言葉」が一時的に出現することがわかった。

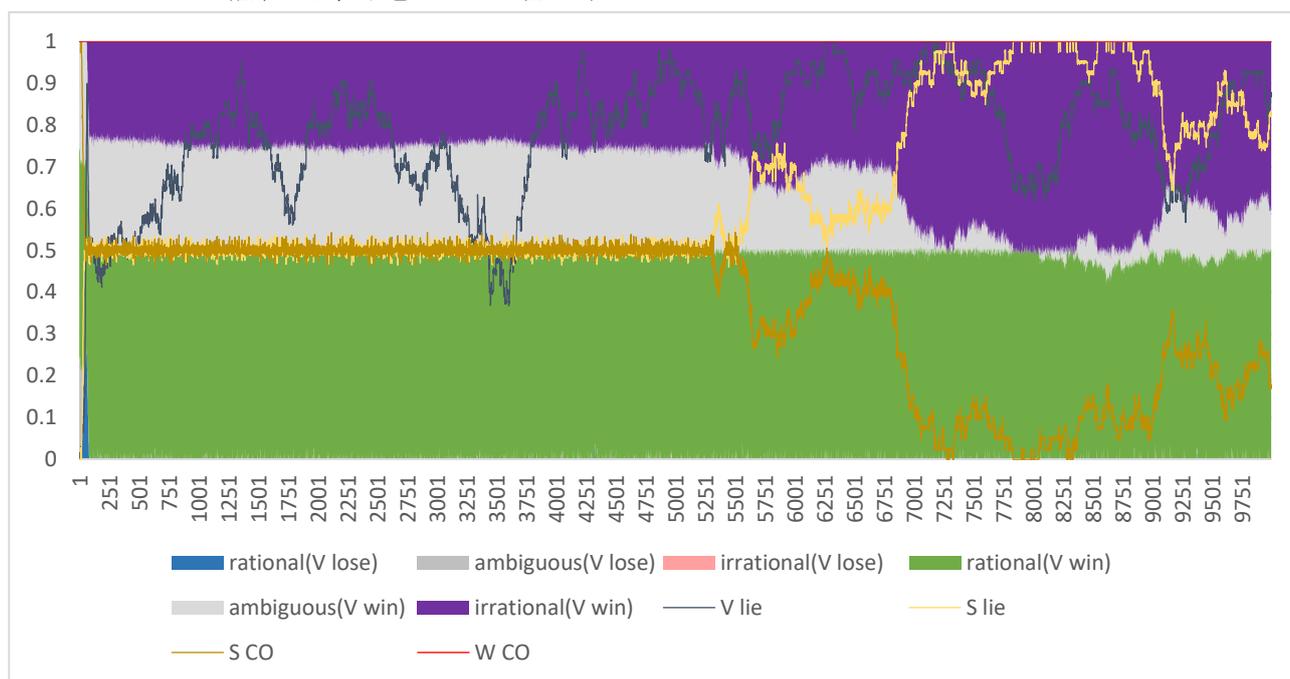


図2：10000世代の村人陣営のみ進化における、村エージェントが用いた戦略と占、狼エージェントの嘘、COの割合（背景の紫の部分は村人陣営が勝利する時に不合理な戦略の使用率）

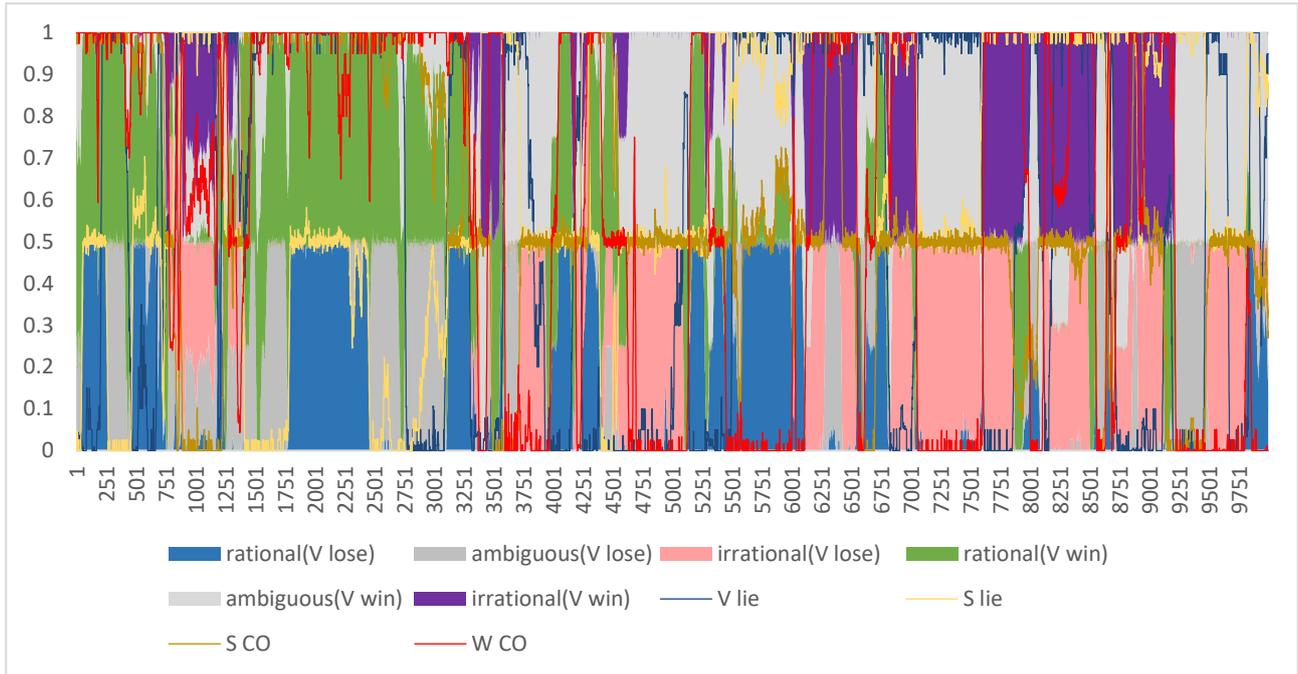


図 2：10000 世代の全エージェント進化における、村エージェントが用いた戦略と占、狼エージェントの嘘、CO の割合（背景の紫の部分は村人陣営が勝利する時に不合理的な戦略の使用率）

### 参考文献：

- [ 1 ] Katagami, D., Takaku S., Inaba, M., Osawa, H., Shinoda, K., Nishino, J., and Toriumi F.: Investigation of the effects of nonverbal information on werewolf. In *IEEE International Conference on Fuzzy Systems*, pp. 982-987, (2014)
- [ 2 ] Bi, X. and Tanaka, T.: Human-side strategies in the werewolf game against the stealth werewolf strategy. In *Lecture Notes in Computer Science*, Vol. 10068, pp. 93-102, (2016)
- [ 3 ] Hung H. and Chittaranian G.: The Idiap Wolf Corpus: Exploring Group Behaviour in a Competitive Role-playing Game. In *Proceedings of the 18<sup>th</sup> ACM International Conference on Multimedia*, pp. 879-882, (2010)
- [ 4 ] 大澤博隆, 汪博豪, 佐藤健: 進化シミュレーションを用いた 3 人狼の分析, 日本ソフトウェア科学会第 34 回大会講演論文集, 2017
- [ 5 ] 大澤博隆, 佐藤健: 3 者間人狼における戦略の検討, 人工知能学会全国大会, 2016, pp. 25-30