

進化シミュレーションを用いたエージェントにおける情報共有の分析

Analysis of Agents' Information Sharing with Evolutionary Computation

汪 博豪*¹ 大澤 博隆*² 佐藤 健*³
 Bohao Wang Hirotaka Osawa Ken Satoh

*¹ 筑波大学大学院システム情報工学研究科
 Graduate School of Systems and Information Engineering, University of Tsukuba

*² 筑波大学システム情報系 *³ 国立情報学研究所
 Faculty of Engineering, Information and Systems, University of Tsukuba National Institute of Informatics

This paper introduced analysis of agents' information sharing with evolutionary computation. We focus on a method that human adds a hidden meaning to our word for sharing information only with cooperators in subgroup. In this paper, we defined 3 player version of werewolf game, and run a 10000 generations evolutionary computation. Then, we classified villager's strategies with label of "rational", "irrational" and "ambiguous". We suggested that when villager used an "irrational" strategy and won the game, seer and villager used the above information sharing method. As the result, the above method appeared in both case of wolf fixed and wolf evolved. In the case of wolf fixed, seer added hidden meaning to a fixed word in 99.2% of times to share information only with the villager. In the case of wolf evolved, in 26.4% of times, seer succeed to share information only with the villager by the above method and seer started to change word with hidden meaning for keeping avoid sharing information with the wolf.

1. はじめに

我々人間の情報共有の手段の一つとして、言語的コミュニケーションは効率的だと考えられる。人間以外の動物と異なり、我々は言語との物に意味を付けて、外部世界や自身の心理活動などの情報を共有する。しかし、言語は話者間で固定的ものではなく、話者間の望みにより移り変わっていく。例えば、あるグループの中で、情報を共有する友人と情報を共有するわけにはいかない他者がいる場合、友人と情報を共有するため、言語の裏の意味を創り出し、友人に裏の意味解釈を誘導させることが起きる。例としては、恥ずかしがりやの人物が好きな人物に「月が綺麗ですね」と言う場合に、実際は「好きです」という裏の意味を伝える例が考えられる。この場合には、好きな人物以外に好きだという原意を知らせたくないため、「月が綺麗ですね」と言うプロセスが発生している。発話者は情報を共有されてない他人に単なる月の話と考えさせ、裏の意味を共有せず、恥ずかしい雰囲気を生み出すことを回避している。

我々は、上記の言語から裏の意味を創り出し、あるグループの中で友人とのサブグループに情報を共有する手法が、エージェント同士のコミュニケーションにおいて発生すると考えている。本研究は進化シミュレーションを用いてエージェントにおける情報を共有する手法を調べた。本研究は人工知能やエージェントなどの様々な研究で用いられている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用いた[Katagami 14][Bi 16][Hung 10]。我々は以前の研究により[大澤 17]、最小系の 3 人狼ゲームを定義し、進化シミュレーションを行った。上記の裏の意味を創り出しの情報共有の手法における分析のため、我々は村人の視点を中心として、協力者/友人の占い師の発話内容が裏の意味を持つかどうかにより、村人と敵対者の人狼の戦略を分類した。占い師の言葉に

裏の意味を持つと考えていた戦略は「不合理的」な戦略と呼び、村人が「不合理的」な戦略を用い、ゲームを勝利した場合は、村人と占い師における言語の裏の意味を創り出し、情報共有の手法が発生したと考えられる。

2. 3 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 人狼ゲーム

上記人狼ゲームの最もシンプルな構成が、3 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話できない。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、村人側はかな

連絡先: 筑波大学大学院システム情報工学研究科, 茨城県つくば市天王台 1-1-1, hailabsec@iit.tsukuba.ac.jp

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、自由会話が行われている。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。

- 「私は村人である」
- 「私は占い師であり、左のプレイヤーが人狼であった」
- 「私は占い師であり、右のプレイヤーが人狼であった」

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[大澤 16]を設定された 3 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発言内容を決める(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。

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。

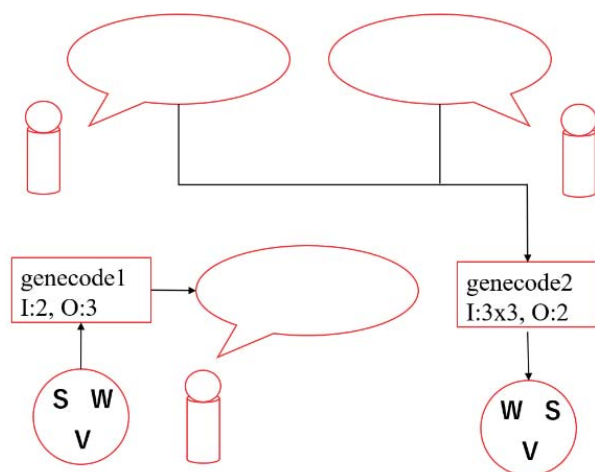


図1: シミュレーションのプロセス

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に考える場合に、合理的な戦略である。それに対し、狼に考える場合に、不合理的な戦略である。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。

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

5. シミュレーションの結果分析

図2の通り、人狼エージェントのみ進化しない場合には、村人と占い師における言語の裏の意味を創り出した情報共有の手法が発生し、99.2%の世代で上記の手法が用いられていた。一方、全エージェントを進化させる場合には、全世代中、上記の手法の発生率は 26.4%となる。

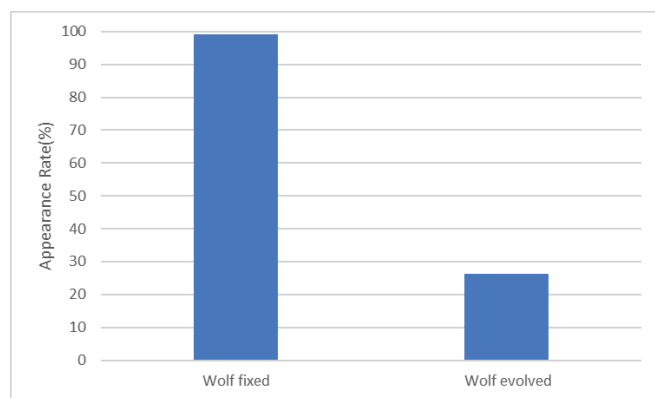


図 2: 全世代における村人が「不合理」な戦略を選択したゲーム勝利の状況の発生率

5.1 考察

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村人と占い師における言語の裏の意味を創り出しの情報共有の手法が発生していた。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た。占い師は「私は村人である」に「私は占い師である」の意味を付け、村人が進化により、占い師の発話の裏の意味を解釈でき、協力者同士に情報を共有できた。人狼エージェントは進化しないため、常に占い師の発話の裏の意味を把握できず、村人と占い師における言語の裏の意味を創り出した情報共有の手法が 99.2%の世代に用いられていた。全エージェントが進化する場合に、狼が占い師の発話の裏の意味を解釈でき始め、村人と占い師のみの裏の情報共有は常に発生することがなかった。故に、26.4%の世代に上記の情報共有の手法が出現していた。狼は占い師と村人の裏の情報共有の手法を把握でき、占い師と同様に村人が騙り始め、占い師と村人の裏の情報共有を破壊した。しかし、占い師も狼に情報を共有しないため、ほかの発話内容に裏の意味を付け始め、再びに村人と情報を共有した。狼はその新たな意味を理解し、占い師はまた新しい意味を創り出すことで、エージェント同士でも人間と同じく、言語が進化する過程が発見できた。

6. おわりに

本研究では 3 人狼ゲームのモデル化を行った後、エージェントにおける協力者同士に言語の裏の意味を創り出す情報共有の手法が出現するかについて、進化シミュレーションの手法を用いて戦略を分類し、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、占い師は常に固定な言葉に裏の意味を創り出し、村人とのサブグループのみに情報を共有した。全エージェントが進化する場合に、占い師は一時的に村人のみと情報を共有していた。占い師は狼に情報を共有しないように新

たな言葉を創り出し、エージェントのコミュニケーションにおいて人間と同じく言語の意味が進化した。

本研究は 3 人グループにおける上記の情報共有の手法について調べた。従って、グループに協力者と敵対者は各一人しかいない条件での分析となる。実世界には、グループ内で協力者と敵対者が複数人いる状況がより一般的だと考えられるため、今後の課題としては、協力者と敵対者が多数の条件で上記の情報共有の手法について調べたいと考えている。

謝辞

本研究は JSPS 科研費 26118006 および公益財団法人 人工知能研究振興財団の助成を受けたものです。

参考文献

- [Katagami 14] 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)
- [Bi 16] 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)
- [Hung 10] Hung H. and Chittaranian G.: The Idiap Wolf Corpus: Exploring Group Behaviour in a Competitive Role-playing Game. In Proceedings of the 18th ACM International Conference on Multimedia, pp. 879-882, (2010)
- [大澤 17] 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
- [大澤 16] 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