

## 強化学習を用いた人狼における最適戦略の抽出

梶原 健吾<sup>†</sup> 鳥海不二夫<sup>‡</sup> 大澤博隆<sup>††</sup> 片上大輔<sup>†‡</sup>

稲葉通将<sup>††</sup> 篠田孝祐<sup>‡‡</sup> 西野順二<sup>†††</sup> 大橋弘忠<sup>‡</sup>

<sup>†</sup> 東京大学工学部システム創成学科 <sup>‡</sup> 東京大学工学系研究科 <sup>††</sup> 筑波大学システム情報系

<sup>†‡</sup> 東京工芸大学工学部コンピュータ応用学科 <sup>‡†</sup> 広島市立大学大学院情報科学研究科

<sup>‡‡</sup> 電気通信大学大学院システム学研究科 <sup>†††</sup> 電気通信大学情報・通信工学専攻

### 1 はじめに

近年、完全情報ゲームであるオセロやチェスに続き、将棋においても人工知能が人間のプロを相手に勝利するようになった。一方で、不完全情報ゲームの人工知能についての研究は少なく発展の余地があり、人工知能が目指すべき新たな目標の一つであると考えられる。

そこでコミュニケーションゲーム“人狼” [1]-[2]に着目し、本研究ではこのゲーム内で使用される主な発話をプロトコル化し、人狼のプラットフォームの構築を行う。また、過去の会話内容やプレイヤー数等を元に、発言内容、能力者の行動等を強化学習し、最適戦略を解析する。

### 2 人狼プロトコル

人工知能が人狼をプレイするためには互いにコミュニケーションを取る必要がある。そこで、人狼をプレイする中で一般的に用いられる発話を基に人狼のプロトコルを作成した。「昨日 A を占ったら、A は人狼だった。」という文章をこのプロトコルを用いて表現したものを図 1 に示す。これは「昨日 A を占った」と「A は人狼である」という二つの事象を、自分の考えを語る意味を持つ“declare”及び確信度を用いてプロトコル言語で表現している。さらに複雑な

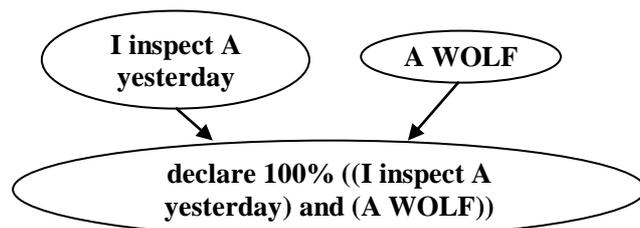


図 1. 発言のプロトコルによる表現

文章も、二文を因果関係で繋ぐ“because”や要求を表す“request”等の言葉を組み合わせて表現可能である。

### 3 人狼の設計

人狼には様々なルールが存在するため、本研究で設計した人狼の設定の概要を以下に記す。

#### 3.1 人数設定

人狼には様々なプレイヤー人数設定があるが、本研究では人狼 BBS<sup>1</sup>で用いている設定を踏襲し、分析を行った。

#### 3.2 会話

本研究では、まず人狼で行われる会話の中で特に重要とされる役職のカミングアウト(以下、CO)、占いの結果、霊媒の結果の 3 つのみを会話として取り入れた。

#### 3.3 役職の CO のタイミング

ゲームを開始する前に各役職の CO する条件を定め、その役職のプレイヤーもしくはその役職を騙っているプレイヤーが CO する条件を満たした際に会話の中で CO する。本研究の実験では占い師と霊媒師は「人狼を見つける」、「自分以外の占い師(霊媒師の場合は自分以外の霊媒師)が出てくる」のいずれかの条件が満たされた時に CO する。

#### 3.4 疑い度合い

各プレイヤーが対象プレイヤーをどの程度疑っているかを表す指標を「個人の疑い度」と

Abstraction of Optimal Strategy in “Are you a Werewolf?” by Reinforcement Learning

Kengo KAJIWARA<sup>†</sup> Fujio TORIUMI<sup>‡</sup> Hirotaka OSAWAWA<sup>††</sup>

Daisuke KATAGAMI<sup>†‡</sup> Michimasa INABA<sup>††</sup>

Kosuke SHINODA<sup>††</sup> Junji NISHINO<sup>†††</sup> Hirotada OHASHI<sup>‡</sup>

<sup>†</sup>School of Engineering, University of Tokyo

<sup>‡</sup>School of Engineering, University of Tokyo

<sup>††</sup>Faculty of Engineering, Information and Systems, University of Tsukuba

<sup>†‡</sup>School of Engineering, Tokyo Polytechnic University

<sup>‡†</sup>Graduate School of Information Sciences, Hiroshima City University

<sup>‡‡</sup>Graduate School of Information Systems, University of Electro-Communications

<sup>†††</sup>Department of Information and Communication Engineering, University of Electro-Communications

<sup>1</sup> <http://www.wolfg.x0.com/>

「各プレイヤー間の繋がり度」の2つに分けた。この2つの指標を組み合わせることで対象プレイヤーの疑い度を評価する。

#### 4 戦略の強化学習

以下に記す学習内容、学習方法で「人狼側のみ学習する」、「人間側のみ学習する」、「双方とも学習する」という3つの試行を1億回ずつ行い、勝率の変化と最適戦略の分析を行う。

##### 4.1 学習内容

###### 4.1.1 投票、占い等の対象選択方法

投票、占い、護衛、襲撃の対象の選択方法を学習する。人狼側は占い(霊能)結果で嘘をつくか否かの選択も学習する。

###### 4.1.2 人狼側が騙る役職

人狼と狂人は村人、占い師、霊媒師の中からどの役職を騙るか学習する。

###### 4.1.3 疑い度合い

「3日目に襲われた」や「AがBを占った」等の状況における「個人の疑い度」と「プレイヤー間の繋がり度」をそれぞれ学習する。

##### 4.2 学習方法

疑い度合いは、特定の状況が当てはまった対象プレイヤーが直近1000回の試行において人狼側であった確率とする。他の項目の学習にはQ学習[3]を用いた。下式がQ学習の式である。

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[ r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

$$\left( \begin{array}{ll} Q: \text{行動価値関数} & \alpha: \text{学習率} \\ s_t: \text{環境} & r: \text{報酬} \\ a_t: \text{行動} & \gamma: \text{割引率} \end{array} \right)$$

戦略の選択にはεグリーディ法[3]を用いた。

#### 5 シミュレーション結果と考察

##### 5.1 結果

各陣営の学習の有無による人間側の勝率の変化を表1に示す。人間側と人狼側の片方だけが学習を行った場合、学習を行った方の勝率が約15ポイント上昇した。双方とも学習した場合は人間側の勝率が2.2ポイント減少した。また人狼側が騙る役職による人間側の勝率の変化を表2に示す。人狼は全員村人として潜伏する時に最も勝率が高くなり、占い師と霊媒師を両方騙った時に最も勝率が低くなった。狂人は占い師を騙るのが最適であった。

##### 5.2 考察

人間側の最適戦略は、占い師は基本的に非CO者の中で疑わしい人を占い、狂人は複数COしている役職の人の中で疑わしくない者を守る

表1. 各陣営の学習の有無における人間側の勝率

	人狼側	学習なし	学習あり
人間側			
学習なし		38.6%	22.3%
学習あり		52.9%	36.4%

表2. 人狼側の騙り役職毎の人間側の勝率

	騙りなし	占い師騙り	霊媒師騙り	両方騙り
人狼	20.9%	36.9%	40.7%	59.2%
狂人	42.0%	32.0%	37.4%	

ことであった。人狼側の最適戦略は、何の役職も騙らない場合は非CO者を襲撃していくことで、占い師を騙る場合は他の占い師が既に占った人を人狼だと占うことであった。序盤にCOした霊媒師は人狼側である確率が高いため、疑わしい霊媒師を吊るすのが人間側の吊りにおける最適戦略となっており、その結果、吊られた人が人間であるか分かりづらくなり、本物の占い師が人間だと占った人を、偽占い師が嘘をついて人狼だと占う際のリスクが減少したと分析できる。

また、生き残り人数が5人または7人の時の襲撃における最適戦略が「誰も襲わない」という一見人狼側に不利な戦略であったが、これはゲームの終盤でプレイヤーの人数を奇数から偶数に調整する戦略であり、実際に上級者によるゲームにおいて人狼側の勝率を上げるために用いられる手法である。

#### 6 結論

本研究では人狼のプラットフォームを構築し、各プレイヤーの戦略の強化学習を行った。その結果、有意な戦略を抽出することが出来た。このシミュレーションの環境では簡単な発言のみでゲームを進行したが、対人戦での人狼では複雑化かつ多様化した発言が出てくるため、今後の課題として、複雑な会話の意味を認識し、その会話に適応した強化学習を可能にすることが挙げられる。

##### 参考文献

- [1] 稲葉通将, 鳥海不二夫, and 高橋健一, “人狼ゲームデータの統計的分析,” in ゲームプログラミングワークショップ2012論文集, 2012, 144-147.
- [2] 大澤博隆, “コミュニケーションゲーム「人狼」におけるエージェント同士の会話プロトコルのモデル化,” in HAIシンポジウム2013, 2013, 122-130.
- [3] 三上貞芳・皆川雅章, 強化学習, 森北出版株式会社, 2000.