

3者間人狼における戦略の検討

Analysis of Strategies in Werewolf Game by Three Players

大澤 博隆^{*1}

Hiroataka Osawa

佐藤 健^{*2}

Ken Satoh

^{*1} 筑波大学

University of Tsukuba

^{*2} 国立情報学研究所

National Institute of Informatics

This study analyzes minimum state of werewolf game that uses three player with roles, villager, werewolf, and seer. In the game, each player tell others once and simultaneously, then the players select a person who is purged simultaneously. The result shows that telling werewolf honestly to other is not always optimal for seer in one-talk game. But if the seer recognize that draw game is lose game, the seer is optimal to say the werewolf position honestly. Under above condition, it is optimal to pretend seer for werewolf. There are no optimal selection for villager.

1. はじめに: 3者間対話における間接的な情報伝達の効果

人間は、3者間の対話では2者間の対話と異なり、自分の対話が間接的に第三者に与える影響を考慮した推論を行って、自らの発言を決定し、エージェントの説得を行っている。例えば2者間の対話では、情報伝達は直接的に行われる。これに対し、3者間の対話では、あるエージェントの発言が自分以外のエージェントにどのような効果をもたらすかを検討し、その結果をもとに自身の行動を決定しなければならない。特に、他者が信頼できる相手かそうでないか不確定な状況下では、こうした間接的な情報伝達の効果を検討することが重要である。

本研究ではこのような3者間対話においてどのような知見が得られるか検討する。題材として、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(片上 et al. 2015)。人狼ゲームに関する研究は昨今盛んになっており、人狼ゲームの理論的分析(Braverman et al. 2008)(西野 2015)(畢 & 田中 2015)や、オンライン人狼ゲームの勝敗データを元にした統計的分析(稲葉 et al. 2012)が行われている。しかし、プレイヤー間に発言が存在する場合の理論的分析の影響はまだ十分に行われておらず、統計的分析から示唆されるような戦略が妥当であるかどうか、理論的な根拠が無い。そのため、本研究では人狼ゲームの最小系である3人狼を考え、どのような戦略がありうるかを検討する。なお、本研究ではとりうる状態を可能世界の組み合わせで表し、確率的な議論を行わない。そのうえで、各プレイヤーがどこまで行動の選択肢を削減できるかを検討する。

2. 人狼ゲームのルール

2.1 人狼ゲームと役職

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。村人陣営はすべての人狼を追放することを目指し、人狼陣営は村人陣営のプレイヤーを襲撃することで数を減らしていく。最終的に以下の勝利条件を満たした陣営が勝利となる。人狼ゲームにおける村人陣営、人狼陣営にはいくつかの特殊能力をもつ

役職が存在する。この中で、最も基本となる役職は、誰が人狼であるかを見分ける占い師である。

ゲームは昼と夜のフェーズに分かれており、昼のフェーズでは対話と追放対象の多数決投票による決定、夜のフェーズでは狼の襲撃先の決定や、占い師の占い先の決定が行われる。人狼が居なくなった場合に村人陣営が勝利となり、人狼の数が人間と同数以上になった場合に、人狼陣営が勝利する。

2.2 3者間人狼の扱う範囲

人狼ゲームは多人数ゲームだが、前項の条件から、2人ではゲームが成立しないため、最少人数は3人となる。本研究では、この3者間人狼ゲームの展開を分析する。3人狼の場合、処刑が一回行われるとゲームが終了するため、夜のフェーズが存在しない。

投票が同数になった場合、人狼ゲームは再投票やランダム投票などの仕組みを用いて投票先を決定し、村人陣営か人狼陣営の勝利を決定する。しかしながら、本研究では簡易化のため、全員が同数投票(3者間人狼では各自1票ずつ投票)された場合、引き分けとして扱う。

3者間人狼で最もシンプルな形式は、村人2人と人狼1人による組み合わせである。ただしこの場合、事前に情報が手に入らないため、村人側が会話によって人狼の場所の情報を得ることができず、そのことが全員に共有されている。よって、情報交換や、その情報交換をかき乱すような偽装の利点がなく、本研究で扱いたいような知的課題に適さない。そのため本研究では、村人の一人を占い師とし、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を開始する。

本研究で仮定する3者間人狼ゲームでは、村人および占い師が村人陣営となり、人狼が人狼陣営となる。

3. 3者間人狼のモデル化

占い師は村人陣営であり、かつ、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、占い師の追放先は人狼であることが最適手となる。従って、本研究で問題となるのは村人、人狼の投票先であり、それがどのように変化するかである。村人側は人狼に投票を2票集めればよいため、村人の選択が人狼投票になる場合に村人側はかならず勝利する。人狼側は自分以外の誰かに2票集めれば勝利する。自分自身に投票できる人間はいないため、人狼の投票先が占

連絡先:大澤博隆
筑波大学システム情報系

い師で、かつ村人の投票が占い師とならない限り、人狼勝利とはならない。これ以外の場合、投票結果は引き分けとなる。

以降、簡易化のためプレイヤーを P_v 、 P_w 、 P_s と呼称する。またプレイヤー P_v を村人、プレイヤー P_w を人狼、プレイヤー P_s を占い師と固定する。もちろんだが、各プレイヤーは他2人が何者であるかは知らない。

3.1 発話のない状況の分類

まず前提として、発話のない状況を考える。発話のない状況での人狼ゲームでは、村人はどちらが狼であるか、という情報を持たない。従って、村人の追放先は占い師もしくは狼側となる。また、人狼は情報を得ることができないため、占い師と村人の判定ができない。従って、こちらも選択肢は削減できず、両方に投票する場合があります。以上より、状態の場合分けの結果は表1のとおりとなる。

表 1: 結果の分類

村人投票先	人狼投票先	結果
人狼	占い師	村側勝利
人狼	村人	村側勝利
占い師	占い師	狼側勝利
占い師	村人	引き分け

3.2 同時一回発話後の投票行動分類

各人が発話し、情報を交換できる場合、それぞれの戦略はどう変わるだろうか。発話の仕方にはいろいろなやり方が考えられるが、本研究では簡易化のため、一度だけ発話ができる、という状況を考える。情報の発話には無限の組み合わせが有りえ、一見情報を伝えないような発話がゲーム内容に影響することも知られている(稲葉 et al. 2013)。しかし、本研究では雑談のように情報を直接伝えない記述は発話とみなさず、情報が伝わる記述を発話とみなす。

発話がない場合には、村人は一切の情報を持たなかったため、前節のようにどちらが人狼であるかという情報をまったく得なかった。これに対し、占い師が発話できる場合には、人狼の場所を占い師が伝えられる可能性が出てくる。ただし、人狼も占い師を騙り、偽の情報を伝えることが可能である。

この条件下で、選択できる発話にはどのような種類が考えられるだろうか。一般的に人狼ゲームとして考えられる発話は、自身の疑い先を示すこと、投票先を示すこと、自身の役職の伝達(カミングアウト/COと称される)、能力結果の伝達、およびその理由説明の5種類が考えられる。前節から、発話のない時点では占い師以外は情報を持たず、占い師もどちらが狼であるか、以外の情報以外を持たないことがわかっている。また、このことは全てのプレイヤー間で共有されている。この状況下で、意味を持つ情報は誰が占い師であるか否かと、その占い結果が何であるかである。

よって、一回発話条件の元で取りうる発話の種類は、自身の役職が占い師であるか否かと、占い師であった時の人狼の場所の伝達を合わせた以下の4通りに限定することができる。発話が一回しかできず、発話が同時に起きる場合、占い師であることを伝えて、狼の場所を伝えない状況よりも、占い師は自身が占い師であること、狼であることを双方を伝えるほうがどの状況でも有利である。なおプレイヤーは3人いるが、自身が占い師でありながら狼と発言することは矛盾するため、その状況はありえない。

• T_0 :「自身が占い師ではない」

- T_v :「自身が占い師であり、狼は P_v (村人)である」
- T_w :「自身が占い師であり、狼は P_w (人狼)である」
- T_s :「自身が占い師であり、狼は P_s (占い師)である」

また、投票行動に関しては、村人、狼それぞれのプレイヤーに2通り(P_v は V_w もしくは V_s 、 P_w は V_v もしくは V_s)の行動が存在する(前ページの説明通り、占い師は狼を吊る行動しか選択肢が無い)。よって、発話行動の組み合わせは $3*3*3=27$ 通り、投票行動は $2*2=4$ 通り、全状態は 108 通りとなる。上記を図1の通り表す。

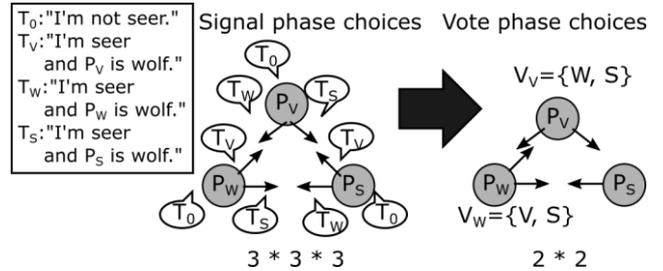


図 1:同時発話のある人狼ゲームの行動分類

3.3 プレイヤーの合理的な行動(相手の投票行動を想定しない場合)

各プレイヤーは自分以外の9通りの発話状態を観察し、その結果としてそれぞれ2通りの投票行動を選択することになる。発話のタイミングは一回しか無いため、各プレイヤーの発話は前の発話に影響を受けない。またそのことをプレイヤー全員が共有しているといえる。

ここで、プレイヤーの行動の合理性を仮定する。ここでは合理的な行動を、考えられる可能な状態の中から、もっともありそうな状態を仮定する行動とする。

(1) 村人の戦略:自身が狼と一人だけから告げられた時、占い師宣言したプレイヤーへの投票

村人プレイヤーにとって、自身を狼という占い師の情報は偽であり、真相を知っていながら嘘をつく占い師であるか、あるいは狼である。この場合、占い師が嘘をつくとは考えづらいため、狼であることを仮定し、自身を狼と言ったプレイヤーに投票することが最善となる。ただし、両方ともが嘘をついていた場合、投票先を決定することはできない。

(2) 村人の戦略:占い結果が一人だけから告げられた時の占い先への投票

村人プレイヤーにとって、自分以外を狼と指すプレイヤーが一人だけいる場合、そのプレイヤーは狼で、占い師が村人を騙ったと考えるより、本物の占い師が狼を指し、狼が村人を騙ったと考えるほうが合理的である。従って、その場合は占い結果が狼と指されたプレイヤーに投票するのが望ましい。ただし、両方ともが占いあった場合、行動を決定することはできない。また、自身が狼と宣言されている場合、そちらへの投票が優先される。

(3) 人狼の戦略:自身が狼と一人だけから告げられた時、占い師宣言したプレイヤーへの投票

人狼プレイヤーにとって、自身を狼と指すプレイヤーが一人だけいた場合、そのプレイヤーが村人でたまたま自分を指したと考えるのは難しく、占い師であることを仮定するのが望ましい。従って、そのプレイヤーに投票することが最善となる。

人狼プレイヤーにとって、自分以外を狼と指すプレイヤーが一人だけいる場合、そのプレイヤーが嘘を付いていることは確実となる。ただし、この場合でも相手が占い師であるか、村人であるかは確定できないため、状態数を削減することはできない。

以上の3戦略を導入した時の、人狼ゲームの結果を分類すると表2のようになる。V_vやV_wがUの場合、どちらとも判別がつかず、可能世界を削減できない状況(両方への投票を含んだ状況)を意味する。結果がV_winのものが村人勝利、結果がW_winのものが狼勝利、drawが引き分けとなる。V_vがWのとき、V_wにかかわらず村人勝利となる。また、V_wがVのとき、狼の勝利はなくなる。結果がcase 3.1となっているのは、村人、人狼ともに情報を持っていないため前節3.1で述べたのと同じ状況であり、村側勝利、狼側勝利、引き分けのいずれをも含んだ状況である。

表2: 一回同時発話の場合の結果の分類 (V_s=W)

P _v	P _w	P _s	V _v	V _w	result
T ₀	T ₀	T ₀	U	U	case 3.1
T ₀	T ₀	T _v	S	U	draw or W_win
T ₀	T ₀	T _w	W	S	V_win
T ₀	T _v	T ₀	W	U	V_win
T ₀	T _v	T _v	U	U	case 3.1
T ₀	T _v	T _w	W	S	V_win
T ₀	T _s	T ₀	S	U	draw or W_win
T ₀	T _s	T _v	S	U	draw or W_win
T ₀	T _s	T _w	U	S	V_win or W_win
T _w	T ₀	T ₀	U	V	V_win or draw
T _w	T ₀	T _v	S	V	draw
T _w	T ₀	T _w	W	U	V_win
T _w	T _v	T ₀	W	V	V_win
T _w	T _v	T _v	U	V	V_win or draw
T _w	T _v	T _w	W	U	V_win
T _w	T _s	T ₀	S	V	draw
T _w	T _s	T _v	S	V	draw
T _w	T _s	T _w	U	U	case 3.1
T _s	T ₀	T ₀	U	U	case 3.1
T _s	T ₀	T _v	S	U	draw or W_win
T _s	T ₀	T _w	W	S	V_win
T _s	T _v	T ₀	W	U	V_win
T _s	T _v	T _v	U	U	case 3.1
T _s	T _v	T _w	W	S	V_win
T _s	T _s	T ₀	S	U	draw or W_win
T _s	T _s	T _v	S	U	draw or W_win
T _s	T _s	T _w	U	S	V_win or W_win

3.4 プレイヤーの合理的な行動(相手の投票行動を仮定した場合の状態削減)

前項の状態が仮定され、全プレイヤーに共有された時の、プレイヤーの合理的な行動を考える。具体的には、各人の取れる発話の選択肢が削減できるかどうかを考える。ある発話を行った時、全ての状態よりも自身の勝利の可能性が増える場合、その戦略はより強い戦略といえ、選択肢を削減できる。表から、村人と人狼の場合、選択肢が削減できないことがわかる。

占い師の選択肢が削減できるかを考える。直感と異なり、占い師が狼の場所を伝えることで、かえって村人陣営敗北の可能性が出る場合が発生することがわかる。占い師の発話には、T₀、T_v、T_wの3通りが考えられる。このうちP_v=T_w、P_w=T_sの場合、占い師が占い師でないと宣言するか、偽の結果を伝える場合には引き分けになるのに対し、占い師が本当の結果を伝えることで、情報が不確定となり、負ける可能性が発生する。

ただし、引き分けを村人陣営の負けと考えた場合、占い師は自身が村人であると嘘をつくよりも、狼の場所を示すことが、必ず強い戦略となる。表2より、T_wの発話を選んだ時、他の2人のプレイヤーがどのような発話を行った場合にも、村人の勝利状況が減る可能性はない。

よって、前項の行動を全てのプレイヤーが仮定しており、かつ村人陣営が引き分けを負けと考えている場合、占い師は必ず狼を当てて告発するのが最適である、と言える。これにより、P_s=T_wを確定でき、表2は表3の通り削減できる。

表3: 一回同時発話の場合の結果の分類の削減 (P_s=T_w, V_s=W)

P _v	P _w	V _v	V _w	result
T ₀	T ₀	W	S	V_win
T ₀	T _v	W	S	V_win
T ₀	T _s	U	S	V_win or W_win
T _w	T ₀	W	U	V_win
T _w	T _v	W	U	V_win
T _w	T _s	U	U	case 3.1
T _s	T ₀	W	S	V_win
T _s	T _v	W	S	V_win
T _s	T _s	U	S	V_win or W_win

次に、表3から狼の発話の選択肢が削減できるかを考える。直感的に考えると、狼は村人を宣言するよりも、占い師を宣言し騙りにいったほうが良いように思える。狼は占い師と異なり、占い師の場所が事前にわからないため、T_vとT_sを自発的に選択することはできず、T₀もしくはT_vとT_sという選択になる。この場合においても、狼勝利の可能性が増えることがわかる。従って、占い師が引き分けを負けと認識して戦略を組んでいる場合には、狼は騙りにいくことが適切であることがわかる。以上より、表4のように削減が可能となる。

表4: 一回同時発話の場合の結果の分類の削減 (P_s=T_w, P_w=T_v or T_s, V_s=W)。結果を並び替えている

P _v	P _w	V _v	V _w	result
T ₀	T _v	W	S	V_win
T _s	T _v	W	S	V_win
T _w	T _v	W	U	V_win
T ₀	T _s	U	S	V_win or W_win
T _s	T _s	U	S	V_win or W_win
T _w	T _s	U	U	case 3.1

今度は村人の発話を考える。村人もどちらが狼であるかの情報を持っていないため、占い師を騙る場合、選べる発話はT₀もしくはT_wとT_sという選択になる。本状況から、可能性を削減することはできないため、村人は騙っても良い、騙らなくても良いことがわかる。

ただし、村人が引き分けを負けと認識する場合には、村人が占い師を騙らないことによって、引き分けの場合を削減できる。

4. まとめ

本研究では、一回発話を行った後に同時投票を行う3者間人狼の分析を行った。

分析結果より、直感に反し、占い師が必ずしも自分の占い結果を告げることが最適とは一般的にいえないことがわかった。ただし、占い師が引き分けを負けと認識している場合には、占い師が人狼の場所を示すことが強い戦略となる。また、その場合には狼はかならず騙ることが最適な戦略となる。

上記の状況下で、村人はどちらの選択をとっても良いが、占い師と同じく引き分けを負けと認識する場合には、騙らずにいることで引き分けの場合を削減できることがわかった。

5. 謝辞

本研究は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

- Braverman, M., Etesami, O. & Mossel, E., 2008. Mafia: A theoretical study of players and coalitions in a partial information environment. *The Annals of Applied Probability*, 18(3), pp. 825-846.
- 稲葉通将 et al., 2013. 雑談ばかりしてると殺される - 人狼BBSにおけるプレイヤーの発言傾向と意思決定・勝敗の分析-. In *JAWS 2013*.
- 稲葉通将, 鳥海不二夫 & 高橋健一, 2012. 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*. pp. 144-147.
- 西野順二, 2015. 自然な人狼の勝率. *研究報告ゲーム情報学 (GI)*, 2015(18), pp. 1-5.
- 畢曉恒 & 田中哲朗, 2015. 対話のない人狼ゲームの戦略. *ゲームプログラミングワークショップ2015論文集*, 2015, pp. 25-30.
- 片上大輔 et al., 2015. 人狼知能プロジェクト. *人工知能学会誌*, 30(1), pp. 65-73.