

戦略記述言語を用いた3者人狼ゲームの解析

大澤博隆^{†1} 佐藤健^{†2}

概要: 本研究では、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、人狼ゲームの最小系である3人人狼を考え、どのような戦略がありうるか検討した。検討のため、各プレイヤーの発言を可能世界の合成として記述するとともに、各プレイヤーのルールを記述するための言語を開発し、総当りで戦略の検討を行った。検討の結果、直感に反し、占い師が積極的に嘘をついた場合でも、村が勝利する場合のあることが判明した。

キーワード: 人狼, シミュレーション, ゲーム

1. はじめに：3者間における人狼

人狼ゲームに関する研究は昨今盛んになっており[1]、人狼ゲームの理論的分析[2][3][4]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[5]が行われている。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。本研究では人狼ゲームの最小系である3人人狼を考え、どのような戦略がありうるか検討する。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検討する。本研究では、以前の研究成果である3者間人狼における発話のモデル[6]を拡張し、可能世界上でありえる状態をルールセットに基づいて検索した。

2. 3者間人狼の定義

2.1 人狼ゲームとは何か

人狼ゲームはスパイ発見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。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者間人狼の定義

3 人人狼の場合、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。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票集めればよいと、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、人狼の投票先が占い師で、かつ村人の投票が占い師とされない限り、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。

3. 3者間人狼のモデル

3.1 記法の定義

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

^{†1} 筑波大学
^{†2} 国立情報学研究所

他 2 人が何者であるかは知らない。

3.2 発話なしの状況分類

前提として、発話のない状況を考えよう。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

3.3 発話のモデル化

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[7]。しかし、本研究では雑談のように情報を直接伝え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何の役職である可能性を持つかと、その占い結果が何であるかである。

本研究では、両記述を合わせ、各プレイヤーの発話を可能世界の集合として記述する。各プレイヤーの役職の割り振りは、 $3P_2$ となり全部で 6 種類となる。一回の発話で、このうちそれぞれの可能性が存在するかしないかを表現する。各プレイヤーごとの発話は、 $2^6=64$ 通りとなる。プレイヤーが 3 人おり、各プレイヤーが自分以外のプレイヤー 2 人のどちらかに投票を最後に行うため、全ての状態数(全状態

空間)は $2^6 * 2^6 * 2^6 * 2^3 = 2,097,152$ 通りとなる。

上記の表現では、例えば、自分が占い師である確率は、6 状態のうち、自分が占い師である 2 状態のみを提案する手法となる。また、人狼を指摘する場合は、あるプレイヤーが人狼である 2 状態のみを提案するやり方となる。自分が CO し、かつ狼を指摘している場合は、可能世界は 1 状態に収束する。

4. 3 者間人狼のモデル

各プレイヤーは自分以外の $64*64$ 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、プレイヤーのルールを記述する。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、両方も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、両方も占いあった場合、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

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。

(4) 占い師の戦略:狼への投票

占い師は狼プレイヤーを知っているため、他のプレイヤーの発言にかかわらず、必ず狼プレイヤーに投票するのが最善である。

以上の4つの戦略をjsonを元にした言語として記述した。この言語を処理するためのプログラムを開発し、探索を行った(最終的なコードを付録A.1に示す)。

5. 3者間人狼の検索結果

4章で説明した4条件のルールを入れた時の、村側勝利、狼側勝利、引き分け、ルール違反の総数はそれぞれ(457373, 228673, 238960, 1172146)、勝利、敗北数を合わせて、その数字で村側勝利数を割った村側勝率は67%となった。また、占い師が明示的にCOするだけだと、村側勝利、狼側勝利、引き分け、ルール違反の総数は(12477, 7881, 8016, 2068778)となり、勝率は61%となって、数字上は下がる。しかし同時に狼を指摘した場合(4352, 256, 16, 2092528)となり、村側勝率は94%まで跳ね上がる。当たり前だが、狼指摘が勝利に大きな貢献を生むことがわかる。

しかし一方で、占い師が自身の結果を言わない(自分が狼や村人の可能性を残す)場合でも、村人が勝利する可能性があることがわかる。占い師が明示的にCOない場合(他の可能性を排除しないもの)を抜き出すと、村側勝利、狼側勝利、引き分け、ルール違反の総数はそれぞれ(436704, 216696, 226848, 1216904)となり、村側勝率は67%と変わらない。また、占い師があえて狼の場所について嘘をつく(COして村人を狼と言う)場合にも、(500, 3904, 4096, 2088652)となって、11%、500通りだが、村側勝利の可能性が残ることがわかった。具体例は、次のとおりである。

* talk

Player 0 (villager) talks "(P0: V, P1: W, P2: S) or (P0: V, P1: S, P2: W) or (P0: W, P1: S, P2: V) or (P0: S, P1: V, P2: W)."

Player 1 (wolf) talks "(P0: W, P1: S, P2: V)."

Player 2 (seer) talks "(P0: W, P1: V, P2: S)."

* vote

Player 0 (villager) voted to Player 1

Player 1 (wolf) voted to Player 2

Player 2 (seer) voted to Player 1

Villagers win.

Wolves lose.

これは、占い師と狼が、両方占い師としてCOし、村人を指して狼と言った場合である。この場合、村人はどちらが偽物か決定できない。よって、狼に投票する場合が上記のように存在する。ただし、狼が占い師を指してCOした場合、村側勝利は存在しない。

6. 結論

本研究では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、人狼ゲームの最小系である3人人狼を考え、どのような戦略がありうるか、戦略記述のための言語を開発して検討した。検討の結果、占い師があえて嘘をついた場合でも、村が勝利する条件があることがわかった。

本研究では動作確認のための、極めて簡単な検証しか行っていない。今後は、各条件の詳細について検討を行っていく予定である。

謝辞 本研究の一部は人工知能研究振興財団、および科研費26118006のサポートを受け行われた。

参考文献

- [1] 片上大輔, 鳥海不二夫, 大澤博隆, 稲葉通将, 篠田孝祐, and 松原仁, “人狼知能プロジェクト,” 人工知能学会誌, vol. 30, no. 1, pp. 65–73, 2015.
- [2] M. Braverman, O. Etesami, and E. Mossel, “Mafia: A theoretical study of players and coalitions in a partial information environment,” *Ann. Appl. Probab.*, vol. 18, no. 3, pp. 825–846, Jun. 2008.
- [3] 西野順二, “自然な人狼の勝率,” *研究報告ゲーム情報学 (GI)*, vol. 2015, no. 18, pp. 1–5, Feb. 2015.
- [4] 畢曉恒 and 田中哲朗, “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*, vol. 2015, pp. 25–30, Oct. 2015.
- [5] 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, vol. 2012, no. 6, pp. 144–147.
- [6] 大澤博隆 and 佐藤健, “3者間人狼における戦略の検討,” in 人工知能学会全国大会, 2016.
- [7] 稲葉通将, 大島菜央実, 鳥海不二夫, and 高橋健一, “雑談ばかりしていると殺される -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*, 2013.

付録

付録A.1 jsonによる3者間人狼のルール記述(占い師がCOし、村人を指した場合でも村人が勝つ条件の記述)

```
{ "rules":  
  [  
    ["TRUE", "村人、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、対狼投票"],  
    ["IF",  
      ["AND",  
        ["INTALK", "wolf", "villager", "wolf"],  
        ["NOT", ["INTALK", "wolf", "villager", "villager"]],  
        ["NOT", ["INTALK", "wolf", "villager", "seer"]],
```

```

["OR",
  ["INTALK", "seer", "villager", "villager"],
  ["INTALK", "seer", "villager", "seer"]
],
["VOTE", "villager", "wolf"]
],
["TRUE", "村人、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、対占い師投票"],
["IF",
  ["AND",
    ["INTALK", "seer", "villager", "wolf"],
    ["NOT", ["INTALK", "seer", "villager", "villager"]],
    ["NOT", ["INTALK", "seer", "villager", "seer"]],
    ["OR",
      ["INTALK", "wolf", "villager", "villager"],
      ["INTALK", "wolf", "villager", "seer"]
    ]
  ],
  ["VOTE", "villager", "seer"]
],
["TRUE", "村人：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、対占い師投票"],
["IF",
  ["AND",
    ["INTALK", "wolf", "seer", "wolf"],
    ["NOT", ["INTALK", "wolf", "seer", "villager"]],
    ["NOT", ["INTALK", "wolf", "seer", "seer"]],
    ["OR",
      ["INTALK", "seer", "wolf", "villager"],
      ["INTALK", "seer", "wolf", "seer"]
    ]
  ],
  ["VOTE", "villager", "seer"]
],
["TRUE", "村人：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、対狼投票"],
["IF",
  ["AND",
    ["INTALK", "seer", "wolf", "wolf"],
    ["NOT", ["INTALK", "seer", "wolf", "villager"]],
    ["NOT", ["INTALK", "seer", "wolf", "seer"]],
    ["OR",
      ["INTALK", "wolf", "seer", "villager"],
      ["INTALK", "wolf", "seer", "seer"]
    ]
  ],
  ["VOTE", "villager", "wolf"]

```

```

],
["TRUE", "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、対村人"],
["IF",
  ["AND",
    ["INTALK", "villager", "wolf", "wolf"],
    ["NOT", ["INTALK", "villager", "wolf", "villager"]],
    ["NOT", ["INTALK", "villager", "wolf", "seer"]],
    ["OR",
      ["INTALK", "seer", "wolf", "villager"],
      ["INTALK", "seer", "wolf", "seer"]
    ]
  ],
  ["VOTE", "wolf", "villager"]
],
["TRUE", "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、対占い師
"],
["IF",
  ["AND",
    ["INTALK", "seer", "wolf", "wolf"],
    ["NOT", ["INTALK", "seer", "wolf", "villager"]],
    ["NOT", ["INTALK", "seer", "wolf", "seer"]],
    ["OR",
      ["INTALK", "villager", "wolf", "villager"],
      ["INTALK", "villager", "wolf", "seer"]
    ]
  ],
  ["VOTE", "wolf", "seer"]
],
["TRUE", "占い師は狼へ投票する"],
["VOTE", "seer", "wolf"],
["TRUE", "占い師が CO する"],
["AND",
  ["INTALK", "seer", "seer", "seer"],
  ["NOT", ["INTALK", "seer", "seer", "villager"]],
  ["NOT", ["INTALK", "seer", "seer", "wolf"]]
],
["TRUE", "占い師が偽狼指摘済み"],
["AND",
  ["INTALK", "seer", "villager", "wolf"],
  ["NOT", ["INTALK", "seer", "villager", "villager"]],
  ["NOT", ["INTALK", "seer", "villager", "seer"]]
],
["TRUE", "村側勝利"],
["WINNER", "villagerside"]
}

```