

ILP を用いた人狼ゲームにおける被襲撃・処刑者の傾向分析

Behaviour analysis of executed and attacked players in Werewolf games by ILP

西崎 絵麻 尾崎 知伸
Ema Nishizaki Tomonobu Ozaki日本大学 文理学部
College of Humanities of Sciences, Nihon University

In Werewolf games, it is important to avoid being attacked and executed to win. In this paper, we apply inductive logic programming to the Werewolf BBS data to derive rich characteristic behavioral structures on executed and attacked players.

1. はじめに

人狼ゲームでより長く生き延びるには、被処刑や被襲撃を逃れることが必要である。本研究では、人狼ゲームを対象にゲーム中のユーザの行動を分析し、実際に処刑、襲撃されたプレイヤー（被害者）の特徴を把握する。具体的には、人狼 BBS を対象に、人狼チームの勝利ゲーム、人間チームの勝利ゲームを各 1 ゲーム、合計 2 ゲームのデータから行動の内容や議論の構造を手作業で抽出し、帰納論理プログラミング (ILP)[1] を用いて分析を行う。被害者を正例、生存者を負例とし、役職の表明や、他のプレイヤーに対しての同意や質問などのゲーム中の行動、議論構造を背景知識とすることで導出された被害者の特徴を報告する。

2. 関連研究

人狼 BBS を対象とした先行研究として、占い師の占い先と処刑対象の投票先に関するログデータを対象に、同調と反駁に着目した分析が行われている [2, 3]。ここで同調とは、他プレイヤーから出された意見や主張に賛同することを表し、反駁とは意見や主張に論じ返すことを表す。分析の結果、人狼ゲームエージェントの戦略構築に有用な知識の獲得を実現している。

先行研究 [4] では、人間らしい戦略を持つエージェントの実現を目的に、人狼 BBS を分析し、エージェントの行動モデルを提案している。プレイヤー自身による役職の表明 (カミングアウト) と能力者による他プレイヤーの役職の判定に関する情報に着目し、人間が任意の局面においてどのような判断を下すのか、その行動選択確率を導出している。

文献 [5] では、社会的相互作用パターンの認識を目的に、人狼を用いた分析を行っている。役職騙りや怪しいふるまいを検出するため、実際の人狼ゲームにおける各プレイヤーの話の長さや、話した回数、妨害された回数、話しているときのピッチの平均と標準偏差など、体の動きや表情、発言情報の分析を行っている。

これらの研究に対し、本研究では先行研究で着目している同調や反駁、役職の表明や能力者による他プレイヤーの役職の判定に加え、他プレイヤーの役職の推測や、質問や回答行動、発言回数を背景知識とし、ILP を用いてルール形式で被害者の特徴を分析する。

3. データセット

3.1 人狼 BBS

本研究では人狼 BBS^{*1} のログデータを利用する。人狼 BBS には、白ログ・赤ログ・青ログ・灰ログと呼ばれる 4 種類のログが存在するが、このうち今回は、全プレイヤーが閲覧・発言 (書き込み) できる白ログを対象にする。対象とするゲームは、人狼勝利のゲーム、人間勝利のゲーム各 1 ゲーム計 2 ゲーム分とする。両ゲームとも、15 人構成 (村人: 8 人, 人狼: 3 人, 占い師: 1 人, 霊能師: 1 人, 騎士: 1 人, 狂人: 1 人) であり、突然死はない。各ゲームのゲーム日数等を表 1 に示す。

表 1: ゲームデータの詳細

	ゲーム 日数	白ログ 総数	1 日あたりの 発言回数の平均
人狼勝利	8	1343	17.04
人間勝利	7	1277	16.25

3.2 事例と背景知識

本研究では、各ゲームの各日における各プレイヤーを事例とし、ゲーム G の Day 日目にプレイヤー A が処刑されたことを `executed(day(G, Day), A)` と表す。同様に、ゲーム G の Day 日目にプレイヤー A が襲撃されたことを `attacked(day(G, Day), A)` と表す。表 2 に使用データの事例数を示す。

表 2: 事例数

	正例	負例
襲撃	10	188
処刑	13	179

各プレイヤーの行動を記述するため、人狼プロトコル [6] で提案されている関係や行動に加え、他プレイヤーに対して質問するなど、20 種類の述語を準備した。これらの述語を利用し、過去の行動やプレイヤー間の関係をルールで表現し、ILP に与える背景知識とする。以下に幾つかの具体例を示す。

役職の表明 「ゲーム G の Day 日目の Prev 日目に、プレイヤー X が役職 R であると表明した」という行動は、下記のルールを用いて表現される。

連絡先: 尾崎 知伸, 日本大学 文理学部 情報科学科, 〒156-8550
東京都世田谷区桜上水 3-25-40, tozaki@chs.nihon-u.ac.jp

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

```

comingout(day(G, Day), Prev, X, R) :-
    prev_days(Prev),
    PDay is Day - Prev,
    comingout(day(G, PDay), X, R).

```

ここで comingout(day(G, PDay), X, R) は、「ゲーム G の PDay 日目にプレイヤー X が役職 R であると表明した」という行動を表す。また prev_days(Prev) は、考慮する過去の範囲（今回は 2 日前まで）を表すルールであり、

```
prev_days(P) :- member(P, [0, 1, 2]).
```

と定義される。

役職の推測 「ゲーム G の Day 日目の Prev 日目に、プレイヤー X がプレイヤー Y の役職を役職 R であると推測した」という行動は、「ゲーム G の Day 日目にプレイヤー X がプレイヤー Y の役職を役職 R であると推測する」という行動を表す述語 estimate(day(G, PDay), X, Y, R) を用い、以下のように定義される。

```

estimate(day(G, Day), Prev, X, Y, R) :-
    prev_days(Prev),
    PDay is Day - Prev,
    estimate(day(G, PDay), X, Y, R).

```

他プレイヤーへの質問 「ゲーム G の Day 日目の Prev 日目に、プレイヤー X がプレイヤー Y に対して質問した」という行動は、「ゲーム G の PDay 日目にプレイヤー X がプレイヤー Y に対して質問した」という行動を表す述語 question(day(G, PDay), X, Y) を用いて、以下のように定義される。

```

question(day(G, Day), Prev, X, Y) :-
    prev_days(Prev),
    PDay is Day - Prev,
    question(day(G, PDay), X, Y).

```

他プレイヤーへの回答 「ゲーム G の Day 日目の Prev 日目に、プレイヤー X がプレイヤー Y に対して回答した」という行動は、「ゲーム G の PDay 日目にプレイヤー X がプレイヤー Y に対して回答した」という行動を表す述語 answer(day(G, PDay), X, Y) を用いて、以下のように定義される。

```

answer(day(G, Day), Prev, X, Y) :-
    prev_days(Prev),
    PDay is Day - Prev,
    answer(day(G, PDay), X, Y).

```

霊能結果の報告 「ゲーム G の Day 日目の Prev 日目に、プレイヤー X がプレイヤー Y の霊能結果 R を報告した」という行動は、「ゲーム G の PDay 日目にプレイヤー X がプレイヤー Y の霊能結果 R を報告した」という行動を表す述語 inquested(day(G, PDay), X, Y, R) を用いて、以下のように定義される。

```

inquested(day(G, Day), Prev, X, Y, R) :-
    prev_days(Prev),
    PDay is Day - Prev,
    inquested(day(G, PDay), X, Y, R).

```

ILP では、事例を表す述語を頭部、背景知識を表す述語の連言を本体部に持つルールを導出する。例えば、「ゲーム G の D 日目において、1 日前（前日）に占い師 (seer) であるとカミングアウトし、0 日前（当日）にプレイヤー Y から質問を受けたプレイヤー X は処刑される」というルールは、役職の表明

(comingout) と他プレイヤーへの質問 (question) を用い、以下のように表現される。

```

executed(day(G, D), X) :-
    comingout(day(G, D), 1, X, seer),
    question(day(G, D), 0, Y, X).

```

4. 学習結果

処刑・襲撃の各データに対し、代表的な ILP システムの一つである Aleph^{*2} を適用し、特徴的なルールの抽出を行った。なお、基本的な集合被覆アルゴリズム (induce) に加え、その変種 (induce_cover) 及び各正例毎のルール導出 (induce_max) の 3 種を用いてルール抽出した。抽出されたルール数は被処刑は 19 ルール、被襲撃は 10 ルールである。

4.1 被処刑者データに対する分析結果

被処刑者に関して抽出されたルールの例を以下に示す。また理解を助けるため、各ルールをグラフで表現しなおしたものを図 1～4 に示す。各頂点・辺のラベルの意味は、図 8 を参照したい。

ルール 1 (図 1):

```

executed(day(A, B), C) :-
    inquested(day(A, B), 0, C, D, wolf),
    estimate(day(A, B), 0, C, E, not(wolf)).

```

被処刑日にプレイヤー C はプレイヤー D の霊能結果が人狼であると発言している。また、同じく被処刑日にプレイヤー C はプレイヤー E の役職を人狼ではないと推測している。霊能師特有の行動をしていることから、プレイヤー C はゲーム中に霊能師とカミングアウトしたことが分かる。人狼ゲームでは、人間チームの戦力となる霊能師を処刑することは考えにくい。このことから、霊能師とカミングアウトしたプレイヤーが 2 人以上現れたと考えられる。本物の霊能師は 1 人しかいないため、人狼チームのプレイヤーが嘘をついていることになる。ここで、霊能師とカミングアウトしているプレイヤーを全員処刑することで、確実に人狼チームのプレイヤーを処刑することが可能となる。このことから、被処刑ルール 1 はゲーム上考えられるルールとなった。

ルール 2 (図 2):

```

executed(day(A, B), C) :-
    estimate(day(A, B), 0, D, C, wolf),
    estimate(day(A, B), 1, E, C, not(wolf)),
    answer(day(A, B), 2, C, E).

```

被処刑日にプレイヤー D はプレイヤー C を人狼であると推測している。また、被処刑日の 1 日前にプレイヤー E はプレイヤー C を人狼ではないと推測している。さらに、被処刑日の 2 日前にプレイヤー C はプレイヤー E へ回答している。ここで、プレイヤー E に関わる行動に着目する。被処刑日の 2 日前に、プレイヤー C がプレイヤー E へ回答していることから、過去にプレイヤー E はプレイヤー C へ質問したと考えられる。さらに、被処刑日の 1 日前にプレイヤー E は、プレイヤー C を人狼ではないと推測していることから、プレイヤー C からの回答の内容を考慮し、人狼ではないと推測したと考えられる。しかし、被処刑日にプレイヤー D からは人狼であると推測されている。人狼であると推測されたことが被処刑の決定打と考えられる。このことから、

*2 <http://www.cs.ox.ac.uk/activities/machinelearning/-Aleph/aleph>

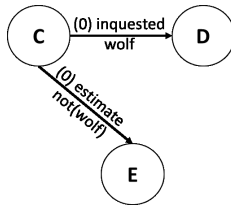


図 1: 被処刑ルール 1

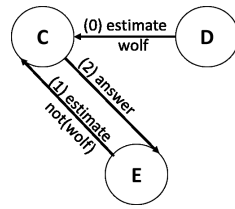


図 2: 被処刑ルール 2

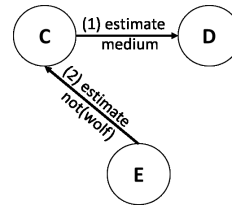


図 3: 被処刑ルール 3

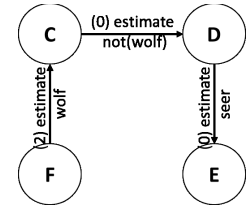


図 4: 被処刑ルール 4

被処刑者を対象とした推測内容は、1 日前の推測内容よりも、当日の推測内容が考慮されるのではないかと考えられる。

このルールを実際のログデータに当てはめると、プレイヤー C は人狼であり、プレイヤー D は狂人もしくは人狼であることが分かった。狂人プレイヤー D が人狼プレイヤー C に対し人狼であると推測した行動は、同じ人狼チームである人狼プレイヤーを処刑に導いてしまう行動である。したがって、メリットのある行動とは言えない。人狼プレイヤー D が人狼プレイヤー C に対し人狼であると推測した行動は、人狼プレイヤー C と仲間と思われると処刑対象になりかねないため、仲間と思われなかったために行動したと考えられる。

ルール 3 (図 3):

```
executed(day(A, B), C) :-
    estimate(day(A, B), 1, C, D, medium),
    estimate(day(A, B), 2, E, C, not(wolf)).
```

被処刑日の 1 日前にプレイヤー C はプレイヤー D を霊能師であると推測している。また、被処刑日の 2 日前にプレイヤー E はプレイヤー C を人狼ではないと推測している。推測内容に霊能師が現れていることに着目すると、霊能師が 2 人以上現れており、本物の霊能師を探すための議論を行っていると考えられる。能力者が 2 人以上現れた場合、誰の能力結果を信じるかという点が、処刑対象者を決める上で重要であると考えられる。したがって、能力者の役職推測は、ゲーム上考えられる議論内容である。

ルール 4 (図 4):

```
executed(day(A, B), C) :-
    estimate(day(A, B), 0, C, D, not(wolf)),
    estimate(day(A, B), 0, D, E, seer),
    estimate(day(A, B), 2, F, C, wolf).
```

被処刑日にプレイヤー C はプレイヤー D を人狼ではないと推測している。被処刑日にプレイヤー D はプレイヤー E を占い師であると推測している。被処刑日の 2 日前にプレイヤー F はプレイヤー C を人狼であると推測している。処刑ルール 3 と同様に推測内容に能力者 (今回は占い師) が現れており、ゲーム上考えられる議論内容であることが分かる。

ルールをログデータに当てはめると、プレイヤー C、プレイヤー E は同一プレイヤーであり、占い師もしくは占い師を騙る人狼プレイヤーであった。このことから、処刑ルール 1 と同様に確実に人狼チームのプレイヤーを処刑するために、占い師と発言しているプレイヤーを処刑していることが分かる。また、占い師と発言しているプレイヤーたちは、自身を占い師と推測しているプレイヤー D を人狼ではないと推測することで、プレイヤー D を味方につけていると捉えることができる。

4.2 被襲撃者データに対する分析結果

被襲撃者に関して抽出されたルールの例を以下に示す。また、各ルールのグラフ表現を図 5~7 に示す。

ルール 1 (図 5):

```
attacked(day(A, B), C) :-
    agree(day(A, B), 2, C, D),
    estimate(day(A, B), 2, C, E, seer),
    question(day(A, B), 1, D, C).
```

被襲撃日の 2 日前にプレイヤー C はプレイヤー D に同意している。被襲撃日の 2 日前にプレイヤー C はプレイヤー E を占い師と推測している。被襲撃日の 1 日前にプレイヤー D はプレイヤー C に質問をしている。このルールをログデータと照合すると、プレイヤー C は村人であり、プレイヤー D は村人、プレイヤー E は占い師もしくは人狼であった。村人プレイヤー C が占い師プレイヤー E を占い師と推測していた場合、プレイヤー C は占い師プレイヤー E の占い結果を信用する可能性が高い。占い結果が人狼の場合、人狼チームにとってプレイヤー C は厄介な人物になると考えられる。このことから、能力者の役職推測を行うプレイヤーを襲撃したと考えられる。

ルール 2 (図 6):

```
attacked(day(A, B), C) :-
    estimate(day(A, B), 0, C, D, wolf),
    question(day(A, B), 2, E, C),
    estimate(day(A, B), 1, E, F, not(medium)).
```

被襲撃日にプレイヤー C はプレイヤー D を人狼と推測している。被襲撃日の 2 日前にプレイヤー E はプレイヤー C に質問している。被襲撃日の 1 日前にプレイヤー E はプレイヤー F を霊能師ではないと推測している。ログデータから、プレイヤー C: 占い師、プレイヤー D: 人狼、プレイヤー E: 騎士、プレイヤー F: 霊能師という結果が得られた。プレイヤー D は占い師を騙った人狼であった。人狼を探す上で、占い師の占い結果は重要であると考えられる。占い師を襲撃することで、占い師の情報 (占い結果) がなくなる。したがって、ゲーム上占い師を襲撃することは考えられるルールである。

ルール 3 (図 7):

```
attacked(day(A, B), C) :-
    estimate(day(A, B), 2, C, D, not(wolf)),
    answer(day(A, B), 0, E, C),
    divined(day(A, B), 2, E, F, not(wolf)).
```

被襲撃日の 2 日前にプレイヤー C はプレイヤー D を人狼ではないと推測している。被襲撃日にプレイヤー E はプレイヤー C に回答している。被襲撃日の 2 日前にプレイヤー E はプレイヤー F の占い結果が人狼ではないと報告している。ログデータより、このルールは、プレイヤー C: 村人、プレイヤー D: 村人、プレイヤー

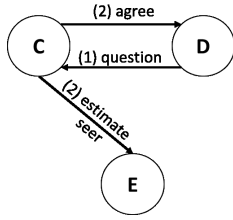


図 5: 被襲撃ルール 1

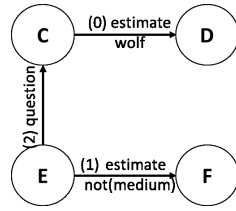


図 6: 被襲撃ルール 2

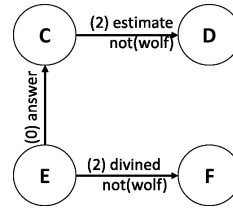


図 7: 被襲撃ルール 3

	プレイヤーCがプレイヤーDに行動
	PDay日前にactionを行動したRは役職を示す ※Rがない行動も存在する

図 8: グラフの見方

E: 占い師もしくは人狼, プレイヤ F: 村人で具体化される. 人狼プレイヤー E は占い師を騙っている人狼である. 被襲撃者であるプレイヤー C について着目すると, 占い師に人狼ではないと占われたプレイヤーであった. 人狼ゲームでは, 占い師に占われ人狼ではないという結果を得たプレイヤーは, 処刑対象から外される場合がある. したがって, このルールでは処刑対象にならないようなプレイヤーを襲撃したと考えられる. しかし, ルール本体部にはプレイヤー F が占われたことを表す述語は現れておらず, 実際に占われたことが襲撃に影響しているかは判断が困難である.

5. 考察

被処刑・被襲撃共に, 約 7 割のルールに *estimate* (推測する) 述語が現れている. さらに, 推測内容は, 複数人現れている能力者 (霊能師や占い師) の役職が顕著である. これより, 能力者に対する推測行動が被処刑・被襲撃に影響を及ぼしていることが分かった.

ルールから, 2 人以上同じ役職を名乗るプレイヤーがいた際, 確実に人狼チームのプレイヤーを処刑するために, 役職を名乗ったプレイヤーを処刑する傾向があることが分かった. 他プレイヤーから人狼であると推測されているプレイヤーもまた, 処刑される傾向があることが確認できた. さらに, 占い師から人狼ではないと判断されたプレイヤーは, (処刑対象とならないため) 人狼に襲撃される傾向も認められた. 実際の人狼ゲームでは, 人狼にとって非常に厄介な役職である騎士が襲撃者候補として上がる. したがって, 人狼は騎士と推測したプレイヤーを襲撃する傾向がある. しかし, 今回の被襲撃ルールからは, 騎士を狙う意図した襲撃であっても, 騎士らしきプレイヤーの特徴が把握できず, 騎士を狙った襲撃か判断することは困難であった. 今回, 能力者を対象とした推測行動が影響を及ぼすことが分かったが, 同じ能力者を名乗るプレイヤーが 2 人以上出現していることが議論を行える最低条件となる. したがって, 能力者と名乗るプレイヤーの人数により, 議論内容も変化すると考えられる.

6. まとめ

本研究では, 人狼ゲームを対象にゲーム中のユーザの行動を分析し, 実際に被処刑者・被襲撃者の特徴の分析を行った. 具体的には, 被処刑者・被襲撃者の行動内容や議論の構造に着目し, ILP を用いて特徴的なルールの抽出を行った. 結果として, 能力者が複数人名乗り出ている場合, 名乗り出た能力者に対する推測行動が影響していることが分かった.

今後の課題として, 対象データの大規模化と共に, 被処刑者・被襲撃者の特徴抽出のさらなる追及があげられる. その際, 能力者の状況を考慮することを検討している. 今回, ILP で抽出されたルールから, 能力者に対する行動が被処刑や

被襲撃に関わることが分かった. 占い師や霊能師, 騎士は人間チームにとっては重要な存在であり, 人狼チームによっては厄介な存在となるため, 非常に重要な人物であると考えられる. 今回使用したゲームは, 占い師, 霊能師共に 2 人名乗り出ているゲームと, 占い師が 2 名, 霊能師が 1 名名乗り出ているゲームである. 占い師や霊能師の名乗り出る人数から, 議論内容が変化し, ルール結果が異なることが予測される. したがって, 能力者に着目し, 名乗り出ている数や日ごとに生存している数などのゲーム状況を取り入れることで, より詳細かつ高精度なルールの導出が期待できる.

謝辞 本研究を行うにあたり, 人狼 BBS の管理者 ninjin 氏及び, データを提供して下さった東京大学鳥海不二夫准教授に感謝いたします.

参考文献

- [1] 古川 康一, 植野 研, 尾崎 知伸:『帰納論理プログラミング』, 共立出版, (2001)
- [2] 稲葉 通将, 大澤 博隆, 片上 大輔, 篠田 孝祐, 鳥海 不二夫: 議論の構造に着目した人狼ゲームの分析, ゲームプログラミングワークショップ 2014 論文集, pp. 61-66, (2014)
- [3] 稲葉 通将, 鳥海 不二夫, 大澤 博隆, 片上 大輔, 篠田 孝祐, 西野 順二: 同調と反駁に着目した人狼ゲームの分析, 人工知能学会全国大会 JSAI2014, (2013)
- [4] 平田 佑也, 稲葉 通将, 高橋 健一, 鳥海 不二夫, 大澤 博隆, 片上 大輔, 篠田 孝祐: プレイログから獲得した行動選択確率を用いた人狼ゲームのシミュレーション, 人工知能学会全国大会論文集, No.29, pp. 1-4, (2015)
- [5] Chittaranjan, Gokul : Hayley Hung. Are you a werewolf? Detecting deceptive roles and outcome in a conversation role-playing game., *In IEEE International Conference on Acoustics Speech and Signal Processing (ICASSP)*, pp.5334-5337, (2010)
- [6] 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, *HAI シンポジウム 2013*, 122-130, (2013)