

人狼ゲームにおけるベイジアンネットワークを用いた推論モデルの検討

Inference Model with Bayesian Networks in Werewolf Game

福井敬徳^{1*} 岩田員典² 伊藤暢浩¹
Takanori Fukui¹ Kazunori Iwata² Nobuhiro Ito¹

¹ 愛知工業大学

¹ Aichi Institute of Technology

² 愛知大学

² Aichi University

Abstract: In the Werewolf game, players's votes greatly affects winning and losing. We propose a method to analyze utterances of players by using Bayesian Networks and decide voting destination. There are numerous factors that determine the voting destination. Therefore we identify the factor that strongly influences the decision of the voting destination by use of decision tree analysis. In addition, a player's program is create based on the inference model constructed by the proposed method and its effectiveness is verified.

1 はじめに

近年、人狼ゲームをプレイできる人工知能の研究が盛んにおこなわれている。その1つに人狼知能プロジェクトがある [1]。このプロジェクトでは人狼ゲームをプレイする人工知能を「人狼知能」と呼んでいる。

人狼知能プロジェクトとは、人狼知能の構築を目指すプロジェクトである。このプロジェクトでは、人狼ゲームを計算機上でプレイすることができる人狼知能プラットフォームが提供されている。また、人狼知能プラットフォームでは人狼ゲームをプレイする「人狼知能エージェント」の開発環境も備えている。さらに、人狼知能エージェントの強さを競うために人狼知能大会が毎年開催されている。

一般的に、人狼ゲームは、複数人のプレイヤーが対面しておこなうゲームとして知られている。プレイヤーは村人陣営と人狼陣営に分かれ、村人と人狼は互いに陣営のプレイヤーの排除を目指し、最終的に残った陣営が勝利する。

また、ゲームが開始するときひとりのプレイヤーに1つずつ、役職が決められる。プレイヤーは割り当てられた役職によって特別な行動・能力を行使することができる。役職については2.2節にて詳細に述べる。

ゲーム上では昼のフェーズと夜のフェーズがあり、2つのフェーズを合わせて1日とする。村人側は昼のフェ

ーズ中にプレイヤー同士で話し合い、投票という行為によって人狼と思われるプレイヤーをゲームから排除する。人狼側は夜のフェーズ中に襲撃という行為によって村人をゲームから排除する。

人狼側も村人と偽って投票をおこなうため、両陣営にとって投票は非常に重要な行動のひとつである。そして投票先の決定は勝率にも大きく影響を与える。

村人側は人狼だと思われるプレイヤーに投票したいため会話を通して正しい投票先を探り出す必要がある。反対に人狼側は、他のプレイヤーの会話から投票されそうなプレイヤーを見つけ出し、そのプレイヤーに投票先を誘導する必要がある。しかし、会話内容は多岐にわたるため、人狼知能エージェントが全ての会話を理解し、どの投票先に投票すべきか判断することは難しい。

会話から投票先決定に必要な情報を得る関連研究として、大川ら [2] の研究がある。大川らの研究では、多層パーセプトロンによる役職推定をおこなっている。過去の人狼知能大会のログから役職推定に有用な発言の特徴を抽出し、多層パーセプトロンに学習させている。これにより、役職を会話から推測し投票先の決定をおこなっている。また、多層パーセプトロンを推論モデルとした人狼知能エージェントを作成し、評価実験をおこなっている。

他にも、梶原ら [3] の研究では村人陣営における人狼の推定が勝利に大きく貢献することを統計的に示した後、SVMを用いて人狼の推定をおこなっている。入力ベクトルにはゲーム内の日数に加え、人狼の推定に有

*連絡先： 愛知工業大学大学院
愛知県豊田市八草町八千草 1247
E-mail: itfukui0922@icloud.com

効であると考えられる5つの情報を入力ベクトルの特徴に加えている。人狼知能大会のログから抽出した特徴量を学習データと評価データに分けて実験したところ、第1回人狼知能大会決勝進出エージェントよりも高い精度で推定可能であることを示した。最後に、学習したSVMを取り込んだ人狼知能エージェントを開発し、ゲームに参加させることで、対戦でも勝率の向上に貢献できることを示した。

これらの研究では、役職推定に焦点を絞っており、投票先の決定には不十分である。自分が人狼の場合、村人陣営の占い師という役職に投票をし、ゲームから排除する必要がある。これは、占い師の占いによって、自分が人狼であると知られてしまう可能性があるためである。しかし、他のプレイヤーは占い師の可能性のあるプレイヤーに対して投票しないため、多数決の結果、ゲームからの排除ができない。そのため、占い師の役職が推定できたとしても、投票による排除ができず、勝率に寄与することが出来ない。

このように、投票先の決定をするためには、役職推定した上で、どこに投票すると勝率が向上するかを会話の流れから考える必要がある。しかし、人狼知能に関する研究の中で投票行為に着目した研究はない。

そのため、本研究では役職推定ではなく、投票行為に着目する。また、会話に基づいた投票を行うために、ベイジアンネットワーク [4] を用いて投票先の決定をする推論モデルを提案する。この推論モデルは人狼陣営にも貢献できるように、人狼側にも適用できるように設計した。

ベイジアンネットワークには2018年8月にCEDEC 2018にて開催された人狼知能大会のログから抽出した要素を取り入れる。また、要素の選出には、決定木分析を用いる。これによって、適切な投票を行えると考えられる。

推論モデルの構築後、このモデルを組み込んだ人狼知能エージェントを作成する。

さらに、評価実験として、同大会の参加エージェントと対戦し、対戦での勝率への貢献を検証する。また、対戦ログから会話に基づいた投票がおこなわれているかを確認する。

2 人狼知能プロジェクト

2.1 人狼知能プロジェクトとは

人狼知能プロジェクトとは、人狼ゲームをプレイする人工知能の構築を目指すプロジェクトである。同プロジェクトは高度な知能の創出、および人と人工知能との高度なコミュニケーションを実現するために、人と自然なコミュニケーションをとりながら人狼ゲームを楽しむことができる人狼知能の構築を目指している。

2.2 人狼ゲーム

人狼ゲームとは、アメリカのゲームメーカーLonny Labs.が2001年に発売したパーティゲーム「汝は人狼なりや？」[5]およびその派生ゲームの総称である。

一般的に人狼ゲームとは、複数人のプレイヤーが対面しておこなうゲームとして知られている。プレイヤーは村人陣営と人狼陣営に分かれ、村人と人狼は互いの排除を目指す。村人陣営のプレイヤーは誰が人狼であるかを知らない。そのため会話によって正体を探る。一方、人狼役職プレイヤーは他のプレイヤーの中で誰が人狼であるかを知っている。

ゲーム上では昼のフェーズと夜のフェーズがあり、2つのフェーズを合わせて1日とする。村人陣営か人狼陣営のどちらかが勝利するまで何日間もおこなわれる(図1)。

昼のフェーズでは、会話によって人狼だと思われるプレイヤーを投票により1名決定し、ゲームから排除する。この排除のことを追放と呼ぶ。夜のフェーズでは、人狼が村人から1名決定し、ゲームから排除する。この排除のことを襲撃と呼ぶ。また夜のフェーズでは、役職によって決められた能力を行使できる。

村人陣営の勝利条件はゲーム上から人狼を追放することである。反対に、人狼の数と人間の数が同数となった場合、人狼陣営の勝利となる。どちらかの陣営の勝利条件を満たすことで、ゲームは終了する。

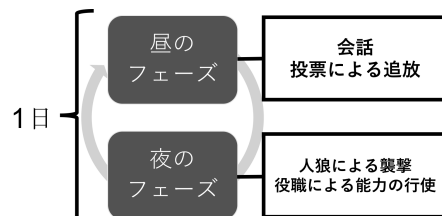


図1: 昼のフェーズと夜のフェーズ

ゲームを開始する時にひとりのプレイヤーに一つずつ、役職が決められる。以下にその役職と役職が持つ特殊な行動・能力について述べる。

村人陣営

村人 : 特別な能力のない役職である。

占い師 : 1日のうちに1人だけ相手が狼であるかを知ることができる。

人狼陣営

裏切り者： 占い師や霊能者の結果からは人間だと判定されるが、この役職は人狼側の勝利条件で自分自身も勝利する。

人狼： 1日のうちに1人だけプレイヤを襲撃することができる。また、人狼同士のみのお話をすることができる。

2.3 人狼知能プラットフォーム

人狼知能プラットフォームとは、人狼ゲームを計算機上でプレイすることができる Java のパッケージである。また、人狼知能プラットフォームでは人狼ゲームをプレイする人狼知能エージェントの開発環境も備えている。

人狼知能プラットフォームは、クライアント・サーバモデルとして構成されている。クライアントは人狼知能エージェントの各行動に対する意思決定を行い、サーバは投票や襲撃といったシステム的な処理をおこなう。また、処理の内容をログデータとして出力する。

人狼知能プラットフォームを用いることで、人狼知能エージェント同士のゲームプレイが可能になる。また、人狼知能プロジェクトが開催する人狼知能大会では、実際に人狼知能エージェントの強さを競うために人狼知能プラットフォームが利用される。

ただし、人間同士がおこなう人狼ゲームにおける、プレイヤの行動や発話の全てを、計算機上で表現することは非常に困難である。そのため、人狼知能プラットフォームでは人狼知能プロトコルというプレイヤの行動や発話を抽象化したプロトコルが定義されている。

人狼知能プラットフォームでは、次の手順でゲームが進行する。

1. 参加するエージェントに役職が割り振られる。
2. 人狼知能エージェント同士がお話をこなう。会話はターン制によって進行し、同ターンの発話は同時に発話されたものとなる。また、各エージェントは1ターンに1回、発話することができる。
3. 20 ターンが経過する、または、全エージェントが「OVER」と発言すると会話が終了する。会話終了後、全エージェントによる多数決によって追放するエージェントを1人決定し、追放されたエージェントをゲームから排除する。
4. 人狼役職のエージェントによる多数決によって襲撃するエージェントを1人決定する。襲撃されたエージェントをゲームから排除する。

5. 2~4 を繰り返し、人狼が全員追放された時点で村人の勝利、村人と人狼が同数となった時点で人狼の勝利となる。

2.4 人狼知能大会

人狼知能プロジェクトの目的を達成する1つの手段として、人狼知能エージェントの強さを競う人狼知能大会が毎年開催されている。

また、人狼知能大会では、5体の人狼知能エージェントが参加するゲームと15体の人狼知能エージェントが参加するゲームがある。それぞれ5人狼、15人狼と呼ぶ。

本研究では、15人狼に比べ複雑な推論を必要としない5人狼を対象とした。

2.5 人狼ゲームにおける投票行為

人狼ゲームは投票によってプレイヤを排除していくゲームである。村人側プレイヤは人狼を昼のフェーズ中の会話によって探り出し、投票先を決定する必要がある。人狼側も自分以外のプレイヤのうち、投票されそうなプレイヤに投票を誘導し、自分が投票されないようにする必要がある。このように、投票行為は村人陣営、狼陣営共に重要な行為であると言える。

2.6 5人狼の特徴

本研究では5人狼を対象とする。5人狼での役職配役数は村人2人、占い師1人、裏切り者1人、人狼1人となっている。1日で追放と襲撃により2人ゲームから排除されるため、2日目は3人でゲームが進行しする。よって、ゲームの最大日数は2日となる。もし、1日目で人狼を追放した場合、2日目は開始されず、ゲーム終了となる。

3 ベイジアンネットワークを用いた投票先決定

3.1 本研究の目的

本研究では、人狼知能の実現に向け、2.5節で示した投票行為に着目する。エージェントが適切な投票先の決定をするためには、会話を理解する必要がある。しかし、人狼知能エージェントにとって全ての会話を理解し、どの投票先に投票すべきか判断することは難しい。ここで、会話を理解するとは、会話に応じた行動を決定することである。これは、現在の人狼知能大会では、人狼知能プロトコルを用いて発言の種類を制限しているものの、会話の展開は多岐にわたっているからである。

そこで、本研究では、会話に基づいた投票行為を実現するために、大会のログからベイジアンネットワークを用いて、推論モデルを構築する。

3.2 提案手法

本研究で提案する推論モデルは、参加者エージェントのうち、どのプレイヤーに投票すべきかを決定する。本推論モデルは、人狼知能エージェントが対戦中に投票先を決定する際、それまでの会話を基に投票先を決定する(図2)。

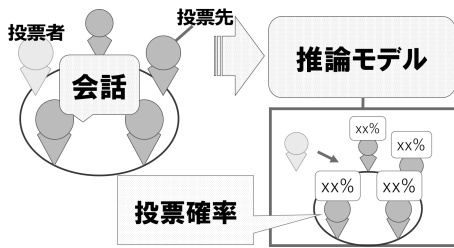


図 2: 推論モデル

推論モデルにはベイジアンネットワークを用いる。ベイジアンネットワークとは、不確実性を含む事象の予測や合理的な意思決定、障害診断などに利用することのできる確率モデルの一種として知られている。そして、ベイジアンネットワークは確率モデルを学習データより学習し、様々な推論をおこなうことができる。本研究では学習データを大会のログから作成した。

4 学習データ

4.1 学習データの作成

本研究では、学習データを人狼知能大会のログから作成し、ベイジアンネットワークを構築する。

学習データの作成は次の手順に沿っておこなう。

1. 人狼知能大会のログから可能な限り会話の要素(観測要素)を抽出する。
2. 投票に強く影響している要素を選出するため、決定木分析をおこなう。
3. 決定木分析結果から選出された要素からベイジアンネットワークの学習データを作成する。

4.2 観測要素

大会のログから抽出できる要素を取り上げ、学習データの候補となる要素を決定する。本研究では抽出した要素を観測要素と定義する。

観測要素とは、会話中に存在する投票先プレイヤーの特徴を表す要素であり、誰に投票するかを決定するのに用いられる。また観測要素は大きく分けて、その投票先が受けている発言、投票先の発言、投票者の発言、ゲームの状況の4つの要素からなる(図3)。これら4つの要素をまとめたものを1レコードとする。例えば1日目の投票者Aのデータセットを考えると、投票先B, C, D, Eの4名が投票先として考えられる。この時、各投票先の特徴を捉えるように観測要素に沿って特徴量をログから抽出する。このため、投票者1人に対して4レコード分のデータが作成される。さらに、投票者をB, C, D, Eとし、同様の操作をおこなう。そのため、1つのゲームログから20レコード分のデータが作成される。

ゲームは1日目で終了する場合と2日目で終了する場合がある。そのため、1日目と2日目は分けて抽出する。さらに、各役職ごとに推論モデルを作成するため、投票者の役職別に抽出する。これは、持ちうる情報の違いと投票先を決める判断基準が異なるためである。そのため、1つのゲームログから抽出された20レコードを4つの各役職別に分ける。

以上より、1日目で各役職ごとに4セット、同様に2日目で4セット抽出する。そのため、1日目で各役職ごとに抽出するため4セット用意する。同様に2日目4セット抽出する。

1日目の観測要素の大別と観測要素数を表1に、2日目の観測要素の大別と観測要素数を表2に示す。

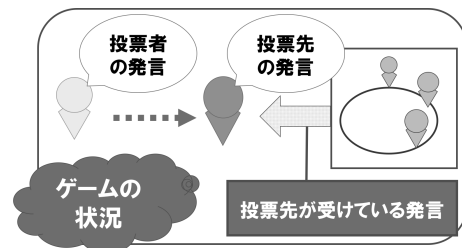


図 3: 観測要素の4つの要素

表 1: 1日目の観測要素数

観測要素の大別	観測要素数
投票先プレイヤーが受けている発言	18
投票先プレイヤーの発言	16
投票者プレイヤーの発言	16
ゲームの状況	1
占い結果 ¹	2

¹ 占い師のみ

表 2: 2 日目の観測要素数

観測要素の大別	観測要素数
1 日目に投票先プレイヤーが受けている発言	18
1 日目の投票先プレイヤーの発言	16
1 日目の投票者プレイヤーの発言	16
1 日目のゲームの状況	1
1 日目の占い結果 ²	2
2 日目に投票先プレイヤーが受けている発言	18
2 日目の投票先プレイヤーの発言	16
2 日目の投票者プレイヤーの発言	16
2 日目のゲームの状況	1
2 日目の占い結果 ¹	2

4.3 決定木分析

人狼知能大会のログから抽出した観測要素は、投票の意思決定と関わりがあるか不明である。そのため、全ての観測要素をベイジアンネットワークの学習データとすることは望ましくない。そこで、投票の意思決定に関わりのある観測要素を見つけるため、決定木分析をおこなう。

決定木分析とは、目的変数をよりよく分けられる説明変数によってグループ分割する手法である。本研究では、決定木分析のアルゴリズムとして Classification and Regression Tree(CART)[6] を用いる。CART 法では目的変数を最も説明する変数をルートノードとして、段階的に分類していく。

決定木分析の目的変数として、その日に投票者が投票した投票先に 1 を、投票していない投票先に 0 を全ての学習データのレコードに対して付加する。この目的変数が付加された学習データを用いて決定木分析をおこなう。決定木分析の結果から、深さの浅い順に観測要素を選出する。本研究では 10 個、15 個、20 個の 3 つにわけ、それぞれベイジアンネットワークを構築し実験をおこなう。そのため、学習データは 8 セット (ゲーム日数×役職数) をそれぞれ 3 つずつ、計 24 セット用意した。

5 ベイジアンネットワーク

5.1 ベイジアンネットワークとは

ベイジアンネットワークの構築には、構築支援ソフトである BayesServer[7] を用いる。ベイジアンネットワークとは、変数間の定量的な関係を条件付き確率で表現した確率モデルである [4]。また、確率変数をノードで表し、変数間の依存関係をグラフ構造で表現する。

変数間の依存関係は、条件付き確率を定義したとき、条件部の確率変数から結果となる変数への向きを持つ有向リンクで表現する。この時、リンクの先に来るノ

ードを子ノード、リンクの元にあるノードを親ノードという。

ベイジアンネットワークの構築には以下の手順に沿って行われる。

1. 決定木分析にて選出された観測要素と決定木分析に用いた目的変数を確率変数として読み込む。
2. 変数間の依存関係を PC アルゴリズム [8, 9] にて求める。
3. PC アルゴリズムによって求められた依存関係を 5.2 節に示す手続きに基づき修正する。
4. 学習データからパラメータ推定をおこなう。

5.2 構造推定

本研究では、因果構造を推定するアルゴリズムとして最も代表的なアルゴリズムである PC アルゴリズムを用いる。PC アルゴリズムは学習データからベイジアンネットワークの構造推定をするため、意味的な因果関係に合うようにリンクを作成できない場合がある。また、学習データの数が十分でない場合、適切な構造推定ができない場合がある。そのため、本研究では次の手順により、構造の修正をおこなう。

1. 占い師は占い結果を取得できる。これは、会話開始前に取得されるため、全ての観測要素の親ノードとなる。このため、リンクの向きが正しくない場合、リンクを反転する。
2. 目的変数のノードは全ての確率変数の子ノードとなる。このためリンクの向きが正しくない場合、リンクを反転する。
3. ベイジアンネットワークを無向グラフとした時、目的変数までのパスがない確率変数は、目的変数へのリンクを作成する。

以上の手順を踏まえ、ベイジアンネットワークの構造を決定する。

6 推論モデルの構築

推論モデルは、参加者エージェントのうち、どのプレイヤーに投票すべきかを決定するモデルである。ベイジアンネットワークで確率推論をおこない、参加プレイヤーへの投票確率を求める。確率推論とは、構築したモデルに確率変数の観測値を入力し、未知のノードの確率分布を求める推論である。この観測値を証拠と呼ぶ。

会話終了後、エージェントによる投票をおこなう直前にそれまでの会話をベイジアンネットワークに証拠

として入れる。その後、投票確率を取得し、最も高い投票確率を得た投票先に投票する。

構築されたベイジアンネットワークを人狼知能エージェントのプログラムに組み込む。投票先を指定する直前に、それまでの会話から観測要素を抽出しベイジアンネットワークに証拠として与える。その後、投票先に投票する確率を求め、最も高い確率を得たプレイヤーに対して投票する。

7 評価実験と考察

7.1 評価手法及び評価対象エージェント

人狼知能大会は人狼知能エージェント同士を対戦させ、勝率を競う大会である。また、大会参加者エージェントは会話から相手の役職を推定し投票先を決定している。

本研究では大会のログから推論モデルを構築した。そのため、会話に基づく投票が可能である上に、ランダムに投票するエージェントに比べて勝率が上がることが期待できる。そこで、本研究で提案した推論モデルが勝率の上昇に貢献しているかを調べるため、推論モデルを組み込んだ人狼知能エージェントを作成する。

実験エージェントは、全部で4つ作成した。会話をしないエージェントとそのエージェントに推論モデルを組み込んだエージェントの2つと、サンプルエージェントとそのエージェントに推論モデルを組み込んだエージェント2つの合計4つである。

NonSpeak

会話をしないエージェント。投票行為はランダムにおこなわれる。

Sample

人狼知能プラットフォームに標準で含まれている適当な会話をするサンプルエージェント。投票行為は組み込まれたサンプルの推論モデルに基づき決定される。

NonSpeak_w/BN

上述の会話をしないエージェントと投票行為以外は同じである。投票行為のみ本推論モデルに従う。

Sample_w/BN

Sample エージェントと会話については同じである。投票行為のみ本推論モデルにより決定する。

人狼知能プラットフォームを利用し、CEDEC2018の大会上位4名と、作成した人狼知能エージェント1名による5人を対戦させ、勝率を調べる。1セット1000回とし、10セット試合をおこなう。その後、「NonSpeak]

と「NonSpeak_w/BN」のウェルチの検定による有意差検定 ($\alpha=0.05$) をおこない、推論モデルによる有意差を確認する。同様に「Sample」と「Sample_w/BN」の有意差検定もおこなう。

7.2 評価結果と考察

勝率を計測した結果を表3に示す。計測した結果、村人陣営の役職は全てにおいて有意差が認められた。反対に、人狼陣営の役職は有意差があるとは言えない結果となった。

また、「Sample_w/BN」の村人役職の観測要素数10と15以外は勝率が向上した。特に「Sample_w/BN」の占い師役職の観測要素数15は26.8ポイントの上昇が確認できた。

対戦結果より、提案手法は村人陣営の勝率には大きく貢献できたが、人狼陣営の勝率は大きく変化しなかった。これは、人狼陣営の裏切り者は人狼の行動に左右されやすく、勝率を上げるためには、投票先決定以外に会話で人狼を助ける必要があるためだと考えられる。同様に、人狼は会話の内容によって投票先を決定するだけでなく、会話によって周りの投票を誘導する必要がある。さらに、襲撃先の決定や裏切り者との協調が必要になるため、提案手法のみでは大きな変化が得られなかったと考えられる。

観測要素数と勝率は正比例の関係にはならないと考えられ、単純に観測要素数を増やしても勝率の向上は期待できない。また、村人の観測要素数15の時に勝率が悪くなった理由として、勝敗に優位な観測要素が抜けていたからだと考えられる。実際にベイジアンネットワークを調べると、目的変数に強く影響を与えている「自身の役職を打ち明けていないプレイヤーから投票を受けているか」の観測要素の有無が人狼推定率に影響を与えているのだと推測できる。ここで、自身の役職を打ち明けることを宣言という。

5人対戦の場合、占い師を宣言しているプレイヤーの多くは本当の占い師か裏切り者であることが多い。自分の役職が村人の場合、占い師を宣言をしていないプレイヤーの2人が投票先の候補となる。

村人役職の対戦ログを解析したところ、占い師を宣言をしていないプレイヤーに対して投票を積極的におこなっていることが判明した。これは、占い師を宣言していないプレイヤーは村人か人狼である可能性が高いため、この宣言をしていないプレイヤーに投票することで人狼を排除できる確率が上がる。つまり、投票先が人狼である確率が4分の1から2分の1に上昇し、それにより勝率が上がったといえる。

また、投票先の決定に際して、役職を宣言していないプレイヤーの投票発言を重視していることもわかった。これは、前述のように占い師を宣言しているプレイヤー

表 3: 人狼知能対戦の 5 人人狼の結果

役職	NonSpeak		要素数	NonSpeak_w/BN		平均勝率の差	有意差 ^a
	勝率平均	標準偏差		勝率平均	標準偏差		
VILLAGER	0.5888	0.0203	10	0.7428	0.0116	0.1540	○
			15	0.7296	0.0183	0.1408	○
			20	0.8374	0.0258	0.2486	○
SEER	0.5684	0.0467	10	0.7840	0.0569	0.2156	○
			15	0.7868	0.0551	0.2184	○
			20	0.8216	0.0137	0.2532	○
POSSESSED	0.0362	0.0094	10	0.0548	0.0086	0.0186	○
			15	0.0664	0.0127	0.0302	○
			20	0.0520	0.0087	0.0158	○
WEREWOLF	0.0772	0.0135	10	0.0882	0.0061	0.0110	○
			15	0.0814	0.0065	0.0042	△
			20	0.0902	0.0076	0.0130	○
役職	Sample		要素数	Sample_w/BN		平均勝率の差	有意差 ^a
	勝率平均	標準偏差		勝率平均	標準偏差		
VILLAGER	0.5776	0.0003	10	0.4762	0.0096	-0.1014	○
			15	0.4558	0.0124	-0.1218	○
			20	0.7054	0.0144	0.1278	○
SEER	0.5598	0.0078	10	0.8206	0.0237	0.2608	○
			15	0.8102	0.0674	0.2504	○
			20	0.8284	0.0286	0.2686	○
POSSESSED	0.0397	0.0001	10	0.0448	0.0129	0.0051	△
			15	0.0678	0.0108	0.0281	○
			20	0.0504	0.0093	0.0107	○
WEREWOLF	0.0926	0.0001	10	0.1320	0.0087	0.0394	○
			15	0.1426	0.0162	0.0500	○
			20	0.1070	0.0347	0.0144	△

^a○: 有意差がある △: 有意差があるとは言えない

は、占い師か裏切り者であり、どちらが信用できるかわからないためである。

一方、人狼役職の対戦ログでは、他の占い師から人狼と占われた投票先は、たとえ占い師を宣言したプレイヤーであっても投票している。人狼は自分1人だけであるため、人狼と占ったプレイヤーは裏切り者であることがわかり、投票先は本当の占い師であることがわかる。そのため、この投票先に投票することは論理的には正しいと言える。しかし、他のプレイヤーは占い師を宣言したプレイヤーに対して、ほとんど投票をしないため、投票による多数決の結果、投票先プレイヤーは追放することができない。このため、勝率に寄与しない。

また、人狼役職の対戦ログの2日目では3人のプレイヤーが残る。この場合、自分以外の2人の中に裏切り者がいる場合、自分が人狼と宣言をした後、裏切り者と票を合わせることで、確実に勝利できる。しかし、提案した推論モデルは投票のみを決定するのみであり、推論モデルによる発言をおこなわない。また、今回作成した推論モデルでは、「投票しようとしているプレイヤーが2日目の裏切り者に投票発言を受けたか」を観測していない。そのため、裏切り者と協調することができず、勝利につながらないと考えられる。

8 おわりに

本研究では、会話の流れからどの投票先に投票すべきかを判断する人狼知能エージェントのための推論モデルを提案した。人狼知能では、人狼知能エージェントが全ての会話を理解し、どの投票先に投票すべきかを判断することは困難である。そこで、提案した推論モデルでは、ベイジアンネットワークを用いて投票先の決定をする。評価実験として、推論モデルを組み込んだ人狼知能エージェントを作成し、CEDEC2018の大会上位4名と対戦させ、勝率への貢献を確認した。その後、対戦ログから会話に基づいた投票がおこなわれているかを確認した。

評価結果より、提案手法は村人陣営の役職の勝率には大きく貢献しており、同時に、会話に基づく投票ができていることを確認した。しかし、人狼陣営の役職に対しては、大きな勝率の変化が見られなかった。これは、村人陣営よりも、人狼陣営は更に複雑な推論を必要とするためだと考える。

本研究の今後の課題として以下の3点を挙げる。

1つ目は観測要素数の選定である。本研究では10, 15, 20の観測要素を選定したが、会話に基づく投票をするために十分な観測要素数であったかを検証する必要がある。特に村人役職の勝率が大きく変化している。この原因となっている観測要素を特定する必要がある。

2つ目はベイジアンネットワークの構造の再検討である。PCアルゴリズムと所定の手続きを定め、ベイジ

アンネットワークの構造を決定した。しかし、より良いアルゴリズムの検討、構造の変化による投票行為への影響など、調査が必要である。

3つ目は推論モデルの応用である。今回作成した推論モデルは、投票先に対して当てはめることで、投票先がどこに投票しようとしているかを推論することが可能だと考えている。これにより、自分自身に投票されないように発言をすることで、相手の投票先を誘導することが可能になると考える。また、観測要素の選択次第では、投票先決定以外にも、人狼の襲撃先決定や、役職推定などにもベイジアンネットワークによる推論が可能であると考えている。これらの手法について検討し、実験をおこないたいと考えている。

謝辞

本研究はJSPS科研費JP16K00310, JP17K00317の助成を受けたものである。

参考文献

- [1] Artificial intelligence based werewolf. <http://aiwolf.org>.
- [2] 大川貴聖, 吉仲 亮, and 篠原 歩. 深層学習を用いて役職推定を行う人狼知能エージェントの開発. *The 22nd Game Programming Workshop 2017*, pages 50–55, 2017.
- [3] 梶原 健吾, 鳥海 不二夫, 稲葉 通将, 大澤 博隆, 片上 大輔, 篠田 孝祐, 松原 仁, and 狩野 芳信. 人狼知能大会における統計分析とsvmを用いた人狼推定を行うエージェントの設計. *The 30th Annual Conference of the Japanese Society for Artificial Intelligence*, 2016.
- [4] 本村 陽一 and 岩崎 弘利. ベイジアンネットワーク技術ユーザ・顧客のモデル化と不確実性推論. 東京電機大学出版局, 2006.
- [5] Are you a werewolf? — looney labs. <http://www.looneylabs.com/games/werewolf>.
- [6] L. Breiman, J. Friedman, C.J. Stone, and R.A. Olshen. *Classification and Regression Trees*. The Wadsworth and Brooks-Cole statistics-probability series. Taylor & Francis, 1984.
- [7] Bayes server. <https://www.bayesserver.com>.
- [8] Judea Pearl 著 and 黒木学 訳. 統計的因果推論—モデル・推論・推測. 共立出版, 2009.
- [9] 斎藤 元幸. 因果性の学習と推論における因果ベイジアンネットワークについて. *認知科学*, 24(1):79–95, 2017.