

# 人狼サーバによる自動対戦を用いた通説の検証 ~人狼は占い師を騙るべきか~

神田直樹<sup>†</sup> 伊藤毅志<sup>†</sup>

**概要:** 人間同士で行う人狼ゲームでは、一種の“通説”と呼ばれるプレイ戦略の指針のようなものによってプレイが行われることがある。しかし、人狼というゲーム自体歴史が浅く、これらの通説には疑わしいものも含まれている。本研究では、人狼知能プロジェクトが提唱する「標準ルール」において、基本的な役職行動と投票行動を行う“基準プレイ”を提案した。そして、人狼知能プロジェクトが提供している人狼サーバの自動対戦機能を用いて、基準プレイにおける勝率と通説を用いた場合の勝率を比較することで、通説の妥当性の検証を試みた。

**キーワード:** 人狼, 通説, 人狼サーバ, 自動対戦

## Verification of Accepted Theories Using Automatic Game Play on Werewolf Server - Should a Werewolf Pretend as a Fortune-teller? -

NAOKI KANDA<sup>†</sup> TAKESHI ITO<sup>†</sup>

**Abstract:** In playing werewolf by humans, they often play based on an "accepted theory" that is a kind of strategic policy. However, history of werewolf game is short, and these accepted theories are doubtful since it has many kinds of local rules. In this research, we proposed a "basic play" that plays basic post behavior and voting behavior on the "standard rule" proposed by werewolf project. We try to verify the validity of the accepted theories by making comparison between the basic play and the play by accepted theories, using automatic game play on werewolf server.

**Keywords:** Werewolf, Accepted theory, Werewolf server, Automatic game play

### 1. はじめに

完全情報ゲームにおける人工知能研究はめざましく、チェスに続き将棋においても、人間トップを超えるレベルになりつつある[1][2]。また、不完全情報ゲームの人工知能の研究も、大貧民やポーカーなどのカードゲームを中心に研究が行われている[3]。一対一のポーカーでは、人間のトップに匹敵するプログラムも出現している[4]。しかし、多人数のポーカーでは、まだ人間のトップには及ばず、研究の余地がある。このようにゲーム AI の研究は、完全情報から不完全情報へ、二人ゲームから多人数ゲームへと広がりを見せている。

近年「人狼」というコミュニケーションゲームがブームになっており、人狼を題材とした TV や動画配信などの番組、小説や舞台なども行われるようになってきている。このブームに合わせて、「人狼知能プロジェクト」というグランドチャレンジが立ち上がり、人狼というゲームを多角的視点から研究しようという動きがある[5]。このプロジェクトでは、人工知能に人狼をプレイさせるための擬似会話プロトコルを備えた人狼サーバを提供しており、このサーバを用いた人工知能による対戦の大会も運営している[6]。

人間同士で行う人狼ゲームでは、一種の“通説”と呼ばれるプレイ戦略の行動指針のようなものが存在している。人狼ゲームはローカルルールが多いばかりか、ゲームとしての歴史も浅く、これらの通説と呼ばれているものの有効性は疑わしい。

そこで本研究では、基本的な役職行動と投票行動を行う

“基準プレイ”を提案し、人狼知能プロジェクトが提供するサーバで用意されている自動対戦可能なシミュレーション機能を用いて、人狼において通説とされているプレイをこの基準プレイに付与することで、実際に勝率が上がるのかを調べ、その通説の有効性を検証していく。

### 2. 人狼とは

#### 2.1 人狼のルール

人狼は、村人と見分けの付かない狼（人狼）がコミュニティに紛れ込んだという設定で行われるコミュニケーションゲームである。ゲームのプレイヤはゲーム開始時に人狼側（人狼、狂人）か村人側（村人、占い師、霊媒師、狩人）の役職に振り分けられる。そして、お互いに自分以外のプレイヤの本当の正体を知らないまま、話し合いの中で相手の正体を探っていく。

村人側に割り当てられたプレイヤは、自分以外のプレイヤがどちらの陣営に属するかは事前には知ることができないが、人狼側のプレイヤは味方の人狼を互いにゲーム開始時に知ることができる。また、村人には聞かれない形で人狼同士の会話は行えるようになっていく。

ゲームは昼と夜の2つのフェーズからなり、昼フェーズでは全てのプレイヤが議論し、人狼とおぼしきプレイヤを投票によって決定し追放する。追放されたプレイヤはプレイから除外される。夜フェーズには、人狼側が相談して人間側のプレイヤから1人を選び襲撃する。襲撃されたプレイヤもプレイから除外される。プレイが進行するにつれて、プレイヤ人数は減っていき、人狼側は人間側の人数を人狼側以下にした時、人間側は人狼側を全滅させた時、それぞれの側が勝利となる。

<sup>†</sup> 電気通信大学情報理工学部研究科情報・通信工学専攻  
Department of Communication Engineering and Informatics, Graduate School of  
Informatics and Engineering, The University of Electro - Communications

村人側には、以下の特殊な能力を持った役職が存在する。「占い師」は、夜フェーズで自分以外の誰かをひとりだけ指名して人狼か否かを占い、正しい結果を知ることが出来る。「霊能者」は、昼フェーズで処刑された人が人狼か否かをその夜に知ることが出来る。「狩人」は、夜フェーズで自分以外のひとりを指名して人狼の襲撃から護衛することが出来る。護衛が成功した場合、その夜は誰も除外されない。

また、人狼の味方をする特殊な役職として「狂人」というプレイヤも存在し、「狂人」は人狼側の勝利が自身の勝利となる。なお、狂人は占い師や霊媒師に占われても村人と判定される。他にも、「共有者」や「妖狐」などプレイ人数が増えた時にのみ現れる特殊な能力を持つ役職も存在する。

人狼知能プロジェクトでは、「人狼標準ルール」が提唱されている[7]。ここでは、上述したような役職やその特殊能力ルール、プレイ人数に応じた役職構成などを公開している。プロジェクトでは、今年7月に CEDEC2015 において、国内初となる人狼知能コンテストを開催した。そこでは、15人のプレイヤからなる村を構成し、それをレギュレーションとした。役職は、構成からなる。村人8人、占い師1人、霊能者1人、狩人1人、人狼3人、狂人1人である。

本研究では、この CEDEC で用いられたルールに準拠して、人狼標準ルールの15人で行う村における通説について考察していく。

## 2.2 人狼の考え方と通説

村人陣営の目的は、人狼陣営（特に騙っている狂人や人狼）の論理的な矛盾を早期に見抜くことであり、人狼を探し出すために占い師と霊媒師による特殊能力によって知り得た情報が重要になる。よって、村人陣営の基本的な戦略的プレイとしては、占い師は早めにカミングアウト（以下CO）して、村人に有益な情報を提供する必要がある。また、霊媒師も適切なタイミングでCOして、村人に有益な情報を提供することが求められる。狩人は潜伏して特殊能力を持つ占い師や霊媒師を守るのが一般的なプレイである。

一方、人狼陣営である狂人は、占い師（稀に霊媒師）を騙り、村人をミスリードするという戦略を取るのが一般的である。人狼は村人のふりをして釣られないように潜伏するか、必要に応じて霊媒師や占い師を騙って、狂人とともに村人のミスリードを誘うプレイを行う。

このようなプレイの戦略的な指針を、本報告では、「通説」と呼ぶことにする。通説には、幾つかの対立する命題も存在する。例えば、「人狼は騙らずに潜伏していたほうが良い」というのと、「人狼は初日から占い師を騙って場を乱したほうが良い」とは相反する命題であり、その戦略の善し悪しについては議論が分かれている。本研究では、これらのプレイヤ間において口伝で伝えられている戦略的指針としての「通説」について、単純なプレイシミュレーションによる勝率を比較することで、その善し悪しについて議論していく。

## 3. 関連研究

人狼の分野において、通説（セオリー）を用いた研究は幾つか行われている。そのうち有名なものとして、Braveman による“Mafia”（人狼とルールがほぼ同様のゲーム題材である）というゲームを用いて、村人、人狼、占い師による確率的統計による各役職の最適戦略を求めた研究がある[8]。これによると、コミュニケーションが全く行われないと、人狼側の勝率が高くなることが示された。

また、西野による人狼陣営が一般に行うと思われる「人狼同士がお互いに投票しない」という自然な戦略を加えた

場合、Braveman の提唱した戦略よりも、人狼陣営の有利さがさらに高くなることを示した[9]。これらの研究では議論を単純化するために村人・人狼のみという条件を前提にしている。しかし、実際の人狼では、他の役職も含まれることが多く、人狼の通説の一般的な研究とは言いがたい。

他のゲームにおいて、通説の真偽を調べる研究としては、とつげき東北氏による麻雀を題材とした研究が挙げられる[10]。とつげき氏は、ネット麻雀に蓄積される膨大な牌譜（ゲームレコード）の中から、統計的な手法で通説について丹念に検証し、幾つかの新説を導いている。この麻雀研究のように、分析可能な十分な量のゲームレコードがある場合はこのような手法は有効であるが、ネット上の人狼である人狼 BBS などではレギュレーションはまちまちで、統計的に分析が行えるだけの十分なデータ量は確保できない。

本研究では、西野の手法を参考にして、人狼の通説の検証を行っていく。これを可能にするために“基準ルール”における各役職のプレイの基準となる行動を確立させる必要がある。そこで本研究では、西野のモデル化を参考に、妥当な人狼モデルに加えて、各々のプレイヤが他のプレイヤが人狼であるかどうかをどれくらい疑っているかという疑い度、役職（村人・占い師・霊媒師・人狼・狂人）の役職に応じた行動を確定して、基準プレイに通説と思われるプレイを変化させることで、その通説の有効性について検証を行っていく。

## 4. 基準プレイ

基準プレイとは、「人狼においてプレイヤ全員が以下のような行動条件の下で、合理的で理知的な行動を取ったと仮定した時のプレイ」と定義する。

### 4.1 発話行動

村人側から見ると、人狼か否かを知りうるプレイヤは占い師と霊媒師しかありえない。基準プレイでは、特定の役職（占い師、霊媒師）をCOしたプレイヤ以外の発言は、プレイに一切影響を与えないものと仮定する。

### 4.2 役職行動

すべてのプレイヤは、以下の役職行動に則って行動するものとする。

【村人】村人からは情報を発信しない。

【占い師】占い師は初日に必ずCO（Coming Out）する。占い師は、一度占った先と他に自称している占い師以外のプレイヤからランダムに1人占う。

【狂人】狂人は初日に必ず「占い師」と騙るものとする。占い師としての占い先は、占い師と同様に占うが、判定結果は以下の確率で黒判定を行い、それ以外の場合は白判定を行うものとする。

$$\text{確率} = \frac{3 - \text{自分が行った黒判定の数} - \text{霊媒師による黒確定の数}}{\text{その日の生存人数}}$$

【人狼】基準プレイでは、何らかの役職を騙らないものとする。襲撃対象は、狩人が自称占い師か霊媒師を守るものと考えて、以下のように決定する。

—確定占い師が存在しない場合：「自称占い師」「霊媒師」「人狼」を除いたプレイヤを対象にランダムで襲撃する。対象プレイヤが存在しない場合は、「自称占い師」「霊媒師」の中からランダムで襲撃する。

—確定占い師が存在した場合：「霊媒師」「白確定」の中からランダムで襲撃する。対象プレイヤが存在しない場合は、「確定占い師（以下の霊媒師の働きによって占い師であることが確定した占い師）」及び「人狼」以外のプレイヤからランダムで襲撃する。但し、生存人数がゲーム開始時の人数半数以下（15人の半数以下は7人以下になった場合）に

なった場合には、「確定占い師」「霊媒師」「白確定」の中からランダムに襲撃する。対象プレイヤーが存在しない場合は、それ以外のプレイヤーからランダムで襲撃する。

**【霊媒師】** 霊媒師は人狼判定が出るまで潜伏し、人狼判定が出たら CO し、それ以降は、霊能結果を毎回報告するものとする。霊媒師の結果と異なる占い結果を告げた占い師は「偽物確定＝狂人」であることが判明する。同時に、もう一人の自称占い師が「確定占い師」となり、その占い師の疑い度は当然白になる。

**【狩人】** 最後まで CO せずに潜伏し、生きている限り、以下の状況に応じて、護衛対象を決定する。

－**確定占い師が存在しない場合**：「自称占い師」「霊媒師」の中からランダムに1人護衛する。ただし、以上の対象が存在していない場合は「白確定」を優先して護衛する。それも存在しない場合は、白に近い対象を護衛する。同色が複数居る場合、ランダムで一人護衛を行う。

－**確定占い師が存在した場合**：「確定占い師」が生き残っている場合には護衛対象は「確定占い師」とする。対象が生存していない場合は、確定占い師が存在しない挙動と同じにする。

### 4.3 投票行動

全てのプレイヤーが理知的なプレイを行っているとは仮定すれば、村人側から見ると、信頼できる役職の情報から明らかな矛盾が生じるまでは、全ての自分以外のプレイヤーは白（完全村人陣営）でも黒（完全人狼陣営）でも無い灰色（村人陣営か人狼陣営か不明）な状態であると考えられる。

会話によって、非論理的な意見の誘導が行われなるとすれば、プレイヤーは占い師・霊媒師の特殊能力にもとづいて得られた知識によって得られた疑い度（村人であれば白、人狼であれば黒であるとして、よりどっちの色に近いかという灰色の色の濃さに対応した疑いの割合）を頼りに投票先を決定するしかない。

よって、本研究で定義する基準プレイでは、自分以外のプレイヤーに対する疑い度に応じて投票先を決定し、その疑い度は占い師・霊媒師の特殊能力による情報によってのみ変更されるものとする。

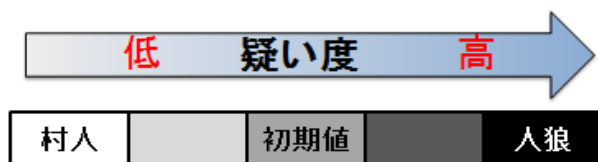


図 1：疑い度の高さを表す灰色の濃さ

占い師が二人 CO した場合、占い師は自分自身の役職を客観的に証明する手段がないため、「自称占い師」という位置づけになる。占い師・霊媒師による特殊能力の情報が無い状態では、村人側のメンバーは、自分以外のすべてのプレイヤーの疑い度の初期値は、図 1 に示すように全員真ん中の灰色からスタートする。占い師の判定結果が村人であった場合は対象プレイヤーの疑い度が薄まるとして、一段階白に近づくものとする。逆に、判定結果が人狼だった場合は対象プレイヤーの疑い度が黒に一段階黒に近づく。もし、二人の「自称占い師」が白と判定すれば、村人確定となり、逆ならば人狼確定となる。このように、すべてのプレイヤーは村人側からみれば、図 1 の 5 段階の疑い度で表現される。

もし、霊媒師の特殊能力により「偽物の占い師」とであると判明した自称占い師は狂人であるとわかり、それ以降の占い結果は、灰色の濃さに影響を与えない。逆に「偽物の占い師」が確定することで、「真の占い師」が判明する。真の占い師が占ってきたこと、今後占うこともすべて信頼され、占われたプレイヤーはすべて白、黒ははっきりする。

“自称占い師”以外の各プレイヤーは、色の濃いプレイヤーから順に投票していく。もし、色の濃いプレイヤーが複数居る場合は、その中からランダムに選択して投票行動を行う。人狼プレイヤーも基本的にこの投票行動に従うが、同じ優先順位のプレイヤーが複数いた場合は、その中から人狼でないプレイヤーに投票するものとする。

## 5. 通説の検証

### 5.1 検証手順

シミュレーションでは、人狼大会[6]で用いられた 15 人のプレイヤーで村を構成し、4 章で挙げた基準プレイにおける人狼側の勝率を計算する。

その後検証したい通説を基準プレイに追加した場合の役職行動と投票行動を検討し、それを満たすプレイをするプレイヤーを構築し、人狼側の勝率を計算する。基準プレイに対して、通説を用いたプレイの勝率がどう変化したかを調べることで、通説の妥当性を検証していく。

本研究では、「通説 A：人狼の 1 人が初日に占い師として騙る」というものと「通説 B：人狼の 1 人が初日占い師として騙り、初日に仲間の人狼を占い白と判定する（それ以降は、通説 A と同じ）」と、村人側が対抗策として占い師ローラーを行った場合「通説 A-R：通説 A で村人側が占い師ローラーを行った場合」「通説 B-R：通説 B で村人側が占い師ローラーを行った場合」の合計 4 つの通説について調べてみることにする。

### 5.2 検証する通説とそれぞれの役職行動と投票行動

#### <通説 A におけるプレイヤー行動>

基準プレイでは、自称占い師は、真の占い師と狂人が騙っている占い師の 2 人しか登場しなかった。しかし、通説 A では、人狼も占い師を騙るので、占い師が 3 名という状況になる。占われたプレイヤーの疑い度の高さは占われた結果と数に応じて、図 2 のように 7 段階に細分化される。

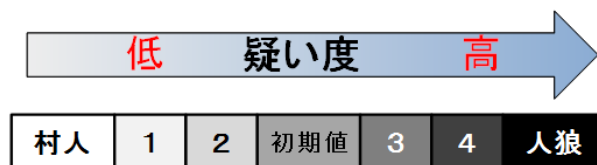


図 2：通説 A における疑い度

各プレイヤーは、上述の疑い度に応じて、投票行動をするものとする。

占い師を騙った人狼は、狂人と全く同じ振る舞いをするものとし、それ以外の役職行動はすべて基準プレイに倣うものとする。

#### <通説 B におけるプレイヤー行動>

通説 B における投票行動は、通説 A と全く同じとする。上述の図 2 の疑い度に応じて、投票行動を行っていく。

役職行動については、占い師を騙った人狼が初日に仲間の人狼のひとりをも「村人」と判定する。それ以外は、狂人と全く同じプレイを続けていく。

#### <占い師ローラーにおけるプレイヤー行動>

上述では通説 A,B で、村人側が何も対策を講じなかった場合であるが、村人側も対策を講じる可能性がある。考えうる対策としては、自称占い師 3 名のうち、少なくとも 2 名は人狼と狂人であるので、自称占い師のローラー（自称占い師を順に追放していく）を行い、霊媒師と連携して、

占い師を騙った人狼をあぶり出すというものである。

もし、占い師ローラー中に、人狼を追放できれば、翌日に霊媒師（もし行かれていれば）によって判明する。そうすれば、人狼が「村人」と占った結果は疑わしくなる。

このように考えると、以下のような投票行動を取ることにする。

プレイヤは、初日に3人の占い師が出てきた場合、その日から、3人のうち誰かにランダムに投票して、追放していく。もし、人狼が追放されたら、霊媒師が名乗りでて、「人狼」であったことが伝えられるので、その結果がわかった瞬間にローラーは終了し、通常の投票行動に移行するものとする。これなら、最悪でも3日目までにひとりの人狼を追放できる。

通説 A,B にローラーを行ったものを、それぞれ「通説 A-R」、「通説 B-R」とする。

### 5.3 シミュレーション結果

基準プレイを1,000回、10,000回、100,000回行った時の人狼側の勝率と通説 A のときの勝率を図3に示す。

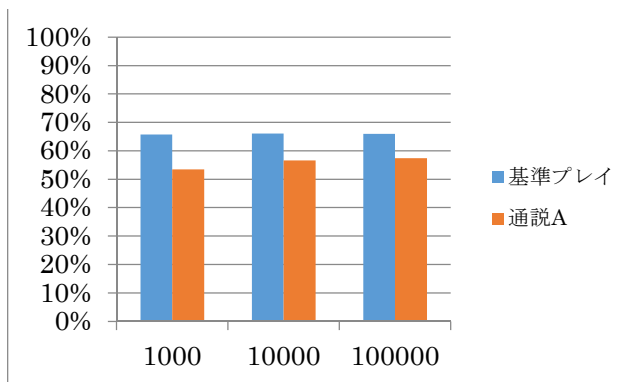


図3: 基準プレイと通説 A の人狼側の勝率

これを見ると、基準プレイでも通説 A でも 1,000 回、10,000 回、100,000 回において、さほど勝率に差がないことがわかる。

また有意差検定を行った所、すべての回数において有意水準 1% で通説 A の方が、有意に勝率が下がったことが示された。

同様に、基準プレイと通説 B の人狼の勝率を比較したところ、図4のような結果が得られた。

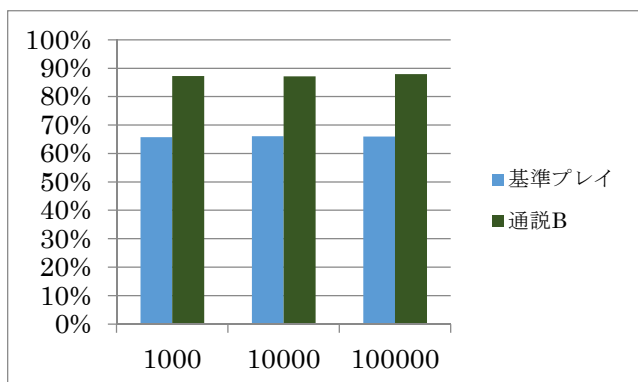


図4: 基準プレイと通説 B の人狼側の勝率

これも同様に有意差検定を行った所、すべての回数において有意水準 1% で通説 B の方が、有意に勝率が上がったことが示された。

さらに、これらのプレイに対して、占い師ローラーを行った「通説 A-R」と「通説 B-R」のプレイについても同様

に人狼側の勝率を調べた所、それぞれ、図5、6のような結果が得られた。

これについても、同様に有意差検定を行った所、すべての回数において有意水準 1% で通説 A-R、通説 B-R の方が、有意に勝率が下がることが示された。

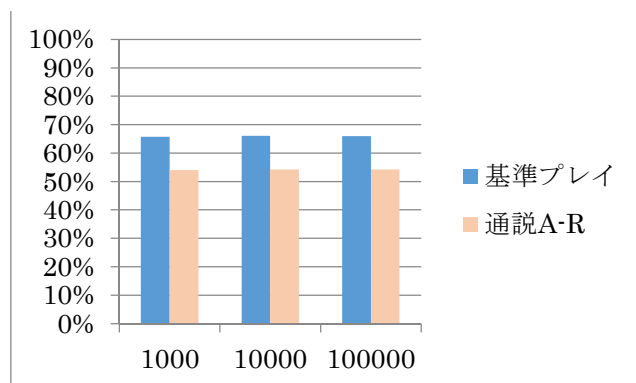


図5: 基準プレイと通説 A-R の人狼側の勝率

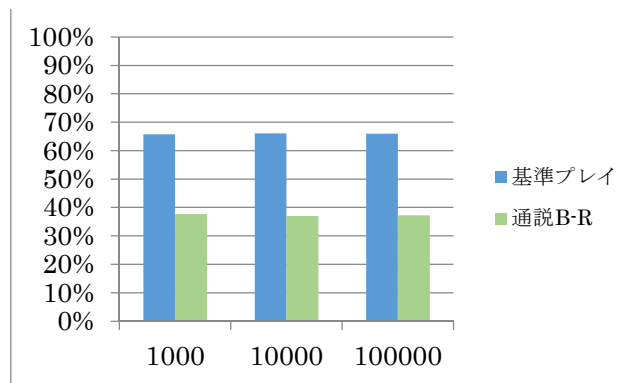


図6: 基準プレイと通説 B-R の人狼側の勝率

### 5.4 結果の考察

図3の結果から、人狼が初日に戦略性なく無作為に占い師を騙ったとしても勝率を上げることは出来ない可能性があることが示唆された。しかし、図4の結果から、人狼が味方の人狼を守るために意図的なプレイをすることで、勝率を上げる可能性があることが示唆された。

さらに、村人側がそれに対抗する「占い師ローラー」を行うことで、人狼側の勝率を下げる事が出来ることも示唆された。

ある一つのプレイを行うと、それに対抗する戦略も生じる。図5,6の結果を見ると、人狼は初日に騙らない方が良いという結論になりそうかと思われるかもしれないが、例えば、占い師ローラーで人狼が霊媒師によって判明してしまいそうなときに、同時にもう一人の人狼が霊媒師を騙って、逆の結果を伝えて場をさらに混乱させるというプレイも考えられる。これを行うことで、人狼側の勝率がどう変わるのかも検証する必要もある。人狼が何日目に発覚するかによって、霊媒師を騙るかどうかを変化させるというような細かな戦略も考えられる。

これらの勝率の比較において、この程度の勝率差を調べる場合、有意水準 1% で検定をするためには、5000 回以上の実験を行えば十分であることが検定力分析の結果わかってきた。今後、この手法で勝率の比較を行う場合には、10,000 回プレイのみで十分に比較が可能であると考えられる。

## 6. おわりに

人狼サーバが提供しているシミュレーションを用いて、プレイヤーの役職行動や投票行動を設定してやることで、「通説」の妥当性の検証に使えることが判明した。

また、通説と考えられているプレイも、それに対抗するプレイを相手陣営が行うことで、勝率が変化することも確認できた。

今後は、さらに様々な通説のバリエーションを調べることで、状況に応じてどのような戦略を取ることが最も良い戦略なのかを考察していきたい。

## 参考文献

- [1] 小谷善行：第3回将棋電王戦を振り返って：3.コンピュータ将棋の棋力の客観的分析－人間のトップに到達したか？－,情報処理 55(8), pp.851-852, (2014).
- [2] 伊藤毅志：人間を超える思考ゲーム AI：－コンピュータ将棋の挑戦と展望－,電気学会誌 134(5), pp.293-296, (2014).
- [3] 伊藤祥平, 但馬康宏, 菊井玄一郎：コンピュータ大貧民における高速な相手モデル作成と精度向上, 研究報告数理モデル化と問題解決 (MPS), 2013-MPS-96(4), pp.1-3, (2013).
- [4] <http://poker.srv.ualberta.ca/about>
- [5] Artificial Intelligence based Werewolf: <http://www.aiwolf.org/>
- [6] 人狼知能コンテスト開催 IN CEDEC2015 : <http://www.aiwolf.org/aiwolf-cedec2015/>
- [7] <http://www.aiwolf.org/howtowagent/>
- [8] Mark Braverman, Omid Etesami, Elchanan Mossel : MAFIA: A THEORETICAL STUDY OF PLAYERS AND COALITIONS IN A PARTIAL INFORMATION ENVIRONMENT, Institute of Mathematical Statistics, 2008 Vol.18, No. 3, 825-846
- [9] 西野順二：自然な人狼の勝率, 情報処理学会研究報告, Vol. 2015-GI-33 No. 18 (2015)
- [10] とつげき東北、伊藤毅志：牌譜の解析による麻雀の分析, 人工知能学会誌 24(3), pp.355-360 (2009).