

役職確定情報に着目した人狼ログ・ダイジェストの作成

Generating a Werewolf Game Log Digest
focusing on Crucial Information on Each Player's Role林 友超*¹ 馬場 瑞穂*¹ 宇津呂 武仁*²
Youchao Lin Mizuho Baba Takehito Utsuro*¹筑波大学大学院システム情報工学研究科
Grad. Sch. Sys. & Inf. Eng. Univ. of Tsukuba*²筑波大学システム情報系
Fclty. Eng. Inf. & Sys. Univ. of Tsukuba

While playing the communication game "Are You a Werewolf", a player always guesses other players' roles through discussions, based on his own role and other players' crucial utterances. The underlying goal of this research is to construct an agent that can analyze the participating players' utterances and play the wolf game as if it is a human. For the first step of this underlying goal, given a specific player participating in the wolf game, this paper studies how to generate a digest of inferring other players' roles from the view of the given specific player. In this inference process, we regard the wolf game rules as well as certain common sense as inference rules. We also show how we examine a real werewolf game log in this framework of generating a digest of inferring the participating players' roles.

1. はじめに

「昼間に人間のふりをして、夜に正体を現して、人間を襲撃する人狼が村に現れた。村人達はお互いを疑いつつ、毎日投票で一番疑わしい一人を処刑する。果たして村人は全ての人狼を排除できるか?」

これが人狼ゲームの簡単なストーリーである。近年、人狼ゲームは日本国内でもよく知られている。例えば、舞台劇中でアドリブで人狼ゲームをプレーする人狼 TLPT*¹や、テレビ番組で芸能人を集めて、人狼ゲームをプレーするもの*²等が知られている。一方、人工知能の研究分野においては、2014年に、人狼ゲームは汎用人工知能の標準問題の一つであるという提案がなされた [篠田 14]。また、2015年8月には、人狼プロジェクト*³が公開したサーバ [鳥海 14]を用いて、第一回の人狼知能大会が開催された*⁴。このように、人狼ゲームを題材とする人工知能研究が近年行われている。ここで、本論文では、人間のレベルの人狼 AI を実現するための基礎的な要素技術の一つとして、人間のプレーヤーによる人狼ゲームのプレーログを自然言語解析し、各プレーヤーの視点において、他プレーヤーの役職が確定していく様子を入狼ログ・ダイジェストとして作成する方式を提案する。この方式においては、各プレーヤーによってなされる多様な役職宣言発言を同定し、この発言内容と人狼ゲームのルール、および、人狼ゲームにおける前提知識として用いられる常識に基づく推論により、他プレーヤーの役職の可能性を絞り込む過程をダイジェストとして可視化する。本論文で利用する役職宣言発言の典型例としては、本物の占い師による自らの役職宣言発言 (占い師 CO)、あるいは、偽りの役職宣言発言によって人狼ゲームを有利に進めようとする人狼・狂人によってなされる偽りの占い師役職宣言発言 (占い師 CO) が挙げられる。今後、人狼 AI を実現する過程においても、本論文で提案する方式と同等レベルの推論に基づいて、他プレーヤーの役職を確定する技術を実現することは不可避である。し

連絡先: 林 友超, 筑波大学大学院システム情報工学研究科,
〒305-8573 茨城県つくば市天王台 1-1-1, 029-853-5427

*¹ <http://7th-castle.com/jinrou/index.php>*² <http://www.fujitv.co.jp/jinroh/index.html>*³ <http://www.aiwolf.org/>*⁴ <http://cedec.cesa.or.jp/2015/session/AC/7649.html>

たがって、本論文の成果は、今後の人狼 AI 研究においても、重要な要素技術の一つとして位置付けることができる*⁵。

2. 人狼

人狼ゲームにおいては、ゲーム開始時に、各プレーヤーが一つの役職を与えられて、役職に応じて「人狼陣営」と「村人陣営」に分けられ、各自の陣営の勝利を目指してプレーする。人狼陣営は、自身の勝利のために、事実と矛盾する発言を行い、村人陣営側を混乱させる。それに対して、村人陣営は、議論の内容の真偽を判別して、人狼の正体を見破り、議論を通じて人狼を処刑することを目指す。人狼ゲームには、何通りかの人数制が設けられているが、一例として、プレーヤー数が16人の場合の役職一覧を表1に示す。

本論文では、ネット上の人狼ゲームサイトである人狼BBS*⁶において公開されているゲームログを対象として、人狼ログ・ダイジェストを作成する。人狼BBSサイトにおいては、各プレーヤーは匿名で人狼ゲームをプレーする。各プレーヤーは掲示板に文字を入力して、議論を行う。

3. 役職特定のための情報

本論文では、計8種類の役職宣言発言 [林 16] を情報源として役職推論を行う*⁷。各発言に加えて、ゲームのルールおよび常識のいくつかを適用することにより、役職候補に関する情報が導出される*⁸。

4. 各プレーヤー視点での人狼ログ・ダイジェスト

本論文の人狼ログ・ダイジェスト作成における役職推論の基本的な考え方は、プレーヤーの多様な役職宣言発言に加えて、ゲームのルールおよび常識のいくつかを適用することにより、

*⁵ [林 16] においては、本研究の枠組みの概略について述べるとともに、プレーヤーへの役職割り当てを行う際に参照するルールおよび常識等について詳細に述べた。一方、本論文においては、役職の割り当てを更新するための詳細な規則群について詳しく述べる。*⁶ <http://ninjin002.x0.com/wolff/>*⁷ 役職宣言発言の典型例としては、役職 CO 宣言が挙げられる。CO とは "Coming Out" の略で、「名乗り出る」という意味。*⁸ 本論文では、人狼における慣習的な設定として以下を仮定する。(i) 本物の占い師、狂人、人狼による占い師 CO、および、本物の霊能者、狂人、人狼による霊能者 CO は、三日目までに行われる。(ii) 本物の狩人、狂人、人狼による狩人 CO は、同一の日にされる。(iii) 共有者 CO は、本物の共有者のみによって行われる。

表 1: 人狼ゲームにおける役職の一覧 (プレイヤー数 16 人の場合)

陣営	生存人数カウントでの扱い	役職	説明	人数
村人	人間	村人	特別な役職を持たない人間である。	7
		占い師	村人陣営の人間である。毎夜、一人を占って、その人が人狼なのか、人間なのかを知り、公表することができる。	1
		霊能者	村人陣営の人間である。投票により処刑されたプレイヤーが、人狼なのか人間なのかを知り、公表することができる。	1
		狩人	村人陣営の人間である。毎夜、一人だけを人狼の襲撃から守ることができる。護衛対象が実際に襲撃をうけたかどうかを知ることができない。	1
		共有者	村人陣営の人間である。もう一人の共有者が誰かを知ることができる。	2
人狼	人狼	狂人	人狼陣営の人間である。人狼の勝利が狂人の勝利となる。狂人と人狼はお互いに正体を知らない。	1
		人狼	毎夜、一人だけ人間を襲撃できる。人狼同士にしか聞こえない会話が可能である。	3
計				16

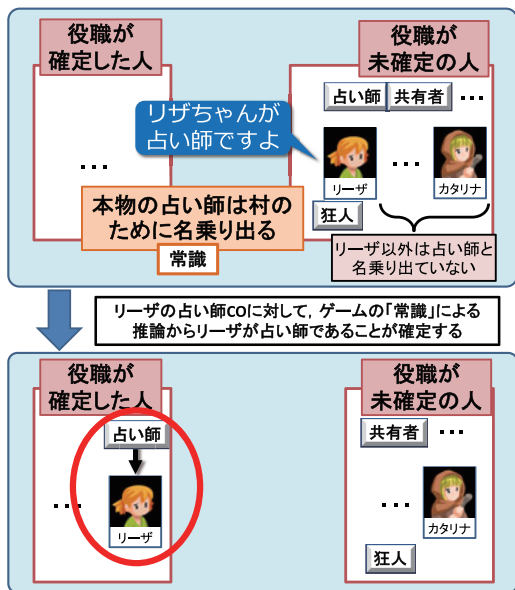


図 1: 役職推論の例

役職推論を行い、各プレイヤーの役職の可能性を絞り込むというものである。図 1 に一例を示す。図 1 では、3 日目までの時点で、リーザだけが占い師 CO し、占い師であると名乗り出ており、一方、他のどのプレイヤーも、占い師であると名乗り出ていない。この時点での、「リーザだけが占い師と名乗り出ている。」という情報と、ゲームの常識「占い師は、村人陣営の勝利のために名乗り出る」、および、ゲームのルール「占い師は一人だけ」の 2 つの知識に基づいて推論を行った結果、占い師候補はリーザ一人だけとなり、リーザが本物の占い師であると確定する。

人狼ログ・ダイジェストのインターフェースのスナップショットを図 2 に示す。この図に示すように、本インターフェースにおいては、特定の日を指定し、特定のプレイヤーの視点のもとで、同日に生存しているプレイヤー（もしくは、同日時点ですでに死亡しているプレイヤー）に対する役職割り当ての（複数の）可能性を提示する。さらに、それらの役職割り当ての根拠となった発言も併せて提示する。

5. 役職割り当ての更新規則

本節では、特定のプレイヤーの視点のもとで、他のプレイヤーに対して役職候補の割り当てを行うとともに、新たな発言が入力されるのに伴って、それらの役職候補の更新を行う際に用い

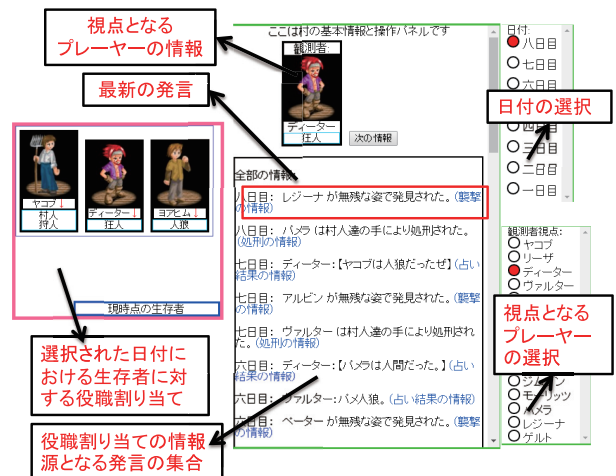


図 2: 人狼ログ・ダイジェストのインターフェース

る規則について述べる。そのための準備として、役職割り当ての対象となるプレイヤーの集合および役職の集合に対する表記法を表 2 に示す。

5.1 発言者・行為者自身に対する役職割り当ての更新規則

プレイヤーが自身に対する役職割り当てに影響を持つ発言または行為を行った場合を対象として、発言者または行為者自身に対する役職割り当てを更新するための規則を表 3 に示す。

5.2 ゲーム中の行為の対象となったプレイヤーに対する役職割り当ての更新規則

ゲーム中にあるプレイヤー X が他のプレイヤー Y に対して、プレイヤー Y に対する役職割り当てに影響を持つ行為を行った場合を対象として、プレイヤー Y に対する役職割り当てを更新するための規則を表 4 に示す。

5.3 役職割り当て更新後の不整合を修正するための更新規則

5.1 節および 5.2 節の規則を用いて役職割り当ての更新を行った結果においては、通常、役職割り当ての候補集合の全体に渡って何らかの不整合が生じる場合が多い。そこで、それらの不整合を修正するための役職割り当て更新規則を表 5 に示す。

6. 役職割り当ての手順

6.1 役職割り当てのアルゴリズム

前節で述べた役職割り当て更新規則を用いることによって、各プレイヤーの役職割り当てを行うアルゴリズムを以下に示す。

表 2: 役職割り当てにおいて用いるプレイヤー及び役職の集合

(a) 役職割り当て $CR(i)$ に依存しないもの

ID	記号	説明
1	RR	ゲームにおける全ての役職の集合
2	$S = \{ \text{占い師, 霊能者, 狩人} \}$	村人陣営において特殊能力を持つ役職の集合
3	$Num(r), r \in RR$	役職 r の総人数
4	P	プレイヤー全体の集合

(b) 役職割り当て $CR(i)$ に依存するもの

ID	記号	説明
1	$CR(i) = \{ \langle p, R \rangle \mid p \in P, R \in RR \text{ はプレイヤー } p \text{ の役職候補集合} \}$	i 日目における全プレイヤーに対する役職割り当ての一つの組み合わせ
2	$CCR(i) = \{ CR(i), \dots, CR'(i) \}$	i 日目における全プレイヤーに対する役職割り当ての組み合わせを全て集めた集合
3	$R(p, i, CR(i))$	i 日目における役職割り当ての組み合わせ $CR(i)$ のもとで、プレイヤー p の役職候補の集合
4	$P(r, CR(i)) = \{ p \in P \mid R(p, i, CR(i)) = \{r\} \}$	i 日目における役職割り当ての組み合わせ $CR(i)$ において、役職が r に確定されたプレイヤーの集合

表 3: 発言者または行為者自身に対する役職割り当ての更新規則

ID	規則
1	プレイヤー p が i 日目に特殊能力を持つ村人陣営の役職 s ($\in S$) を CO した場合は、 i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において、プレイヤー p の役職候補の集合 $R(p, i, CR(i))$ 中の役職として、 s , 狂人, 人狼のみを残す。 $R(p, i, CR(i)) = R(p, i, CR(i)) \cap \{s, \text{狂人}, \text{人狼}\}$
2	プレイヤー p が i 日目に共有者 CO を行った場合は、 i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において、プレイヤー p の役職候補の集合 $R(p, i, CR(i))$ 中の役職として、「共有者」のみを残す。 $R(p, i, CR(i)) = \{\text{共有者}\}$
3	プレイヤー p が i 日目に占い結果を出した場合 (s を「占い師」とする)、または、霊能結果を出した場合 (s を「霊能者」とする) は、 i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ を、役職割り当ての組み合わせ $CR'(i)$ と $CR''(i)$ に分割する。ここで、 $CR'(i)$ においてはプレイヤー p の役職候補を s に確定し、一方、 $CR''(i)$ においてはプレイヤー p の役職候補の集合から s を除外する。 $CR'(i) = (CR(i) - \{ \langle p, * \rangle \}) \cup \{ \langle p, \{s\} \rangle \}$ $(\langle p, R \rangle \in CR(i) \text{ として}) CR''(i) = (CR(i) - \{ \langle p, * \rangle \}) \cup \{ \langle p, R - \{s\} \rangle \}$ $CCR(i) = (CCR(i) - \{CR(i)\}) \cup \{CR'(i), CR''(i)\}$

6.1.1 初期化

$i = 1$ 日目における役職割り当ての組み合わせの初期値として、集合 P 中の全プレイヤー p について、役職候補集合を全役職集合 RR とした役職割り当て $CR_0(i = 1)$ を用いる。

$$CR_0(i = 1) = \{ \langle p_1, RR \rangle, \dots, \langle p_n, RR \rangle \} \quad (\text{ただし, } n = |P|)$$

また、役職割り当て $CR_0(i = 1)$ において、集合 P 中の全プレイヤー p に対する役職割り当ての初期値は全役職集合 RR となる。

$$\forall p \in P, R(p, i = 1, CR_0(i = 1)) = RR$$

そして、役職割り当ての組み合わせの集合 $CCR(i = 1)$ の初期値を次式とする。

$$CCR(i = 1) = \{ CR_0(i = 1) \}$$

6.1.2 i 日目における役職割り当ての更新

i 日目の発言 U および行為 A を全て処理し終えるまで以下の手順を繰り返し実行する。

一つの発言または行為に対する役職割り当ての更新 発言 U

または行為 A を一つ処理し、それが役職宣言発言、占い結果、霊能結果、襲撃情報のいずれかである場合には、表 3 または表 4 における該当規則を適用する。

役職割り当て更新後の不整合の修正 一つの発言または行為に対して役職割り当ての更新を行った後、全てのプレイヤーの役職割り当てが変化しなくなるまで、表 5 中の全ての規則を繰り返し適用する。

6.1.3 日付の更新

i 日目が最終日の場合には終了する。 i 日目が最終日でない場合には、次式によって、役職割り当ての組み合わせの集合の日付の更新を行い、前節の手順によって「 $i + 1$ 日目における役職割り当ての更新」を実行する。

$$CCR(i + 1) = CCR(i)$$

6.2 具体例

ここでは、図 1 を例として、前節で述べた役職割り当てのアルゴリズムの適用過程について説明する。

まず、初期段階におけるリーザに対する役職割り当てとしては、全役職集合 RR が割り当てられる。次に、リーザが占い師 CO した時点において、表 3 の規則 1 が適用され、リーザに対する役職候補の集合は、 $\{ \text{占い師, 狂人, 人狼} \}$ となる。そのまま、3 日目まで、他のどのプレイヤーも占い師 CO しなかった場合には、表 5 の規則 6 が適用され、他の全てのプレイヤーに対する役職割り当てから占い師の割り当てが削除される。最

表 4: ゲーム中の行為の対象となったプレイヤーに対する役職割り当ての更新規則

ID	規則
1	(本物の) 占い師による占い結果, または, 霊能者による霊能結果において, プレイヤー p が「人狼」の場合, i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において, プレイヤー p の役職候補の集合 $R(p, i, CR(i))$ 中の役職として, 「人狼」のみを残す. $R(p, i, CR(i)) = R(p, i, CR(i)) \cap \{\text{人狼}\}$
2	(本物の) 占い師による占い結果, または, 霊能者による霊能結果において, プレイヤー p が「人間」の場合, i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において, プレイヤー p の役職候補の集合 $R(p, i, CR(i))$ 中の役職として, 「人狼」を削除する. $R(p, i, CR(i)) = R(p, i, CR(i)) \cap (RR - \{\text{人狼}\})$
3	プレイヤー p が i 日目に人狼に襲撃された場合, i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において, プレイヤー p の役職候補の集合 $R(p, i, CR(i))$ 中の役職として, 「人狼」を削除する. $R(p, i, CR(i)) = R(p, i, CR(i)) \cap (RR - \{\text{人狼}\})$

表 5: 役職割り当て更新後の不整合を修正するための更新規則

ID	規則
1	i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において, 任意の役職 $r (\in RR)$ について, 役職が r に確定したプレイヤーの人数が, 役職 r の上限人数に一致した場合 ($ P(r, CR(i)) = Num(r)$), 役職が r に確定したプレイヤー以外のプレイヤー $p (\forall p \in (P - P(r, CR(i))))$ に対して, 役職候補の集合 $R(p, i, CR(i))$ 中の役職として, r を削除する. $R(p, i, CR(i)) = R(p, i, CR(i)) \cap (RR - \{r\})$
2	i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において, $CCR(i)$ の削除の必要性のうちの以下の可能性を検証し, 削除が必要な $CR(i)$ を削除する. 具体的には, $CCR(i)$ 中で, 少なくとも一つの役職 $r (\in RR)$ について, 役職候補として r を割り当てられたプレイヤーが一人もいない場合 ($\forall \{p, R\} \in CCR(i), r \notin R$), $CCR(i)$ から $CR(i)$ を削除する. $CCR(i) = CCR(i) - \{CR(i)\}$
3	i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において, $CCR(i)$ の削除の必要性のうちの以下の可能性を検証し, 削除が必要な $CR(i)$ を削除する. 具体的には, $CCR(i)$ 中で, 少なくとも一人のプレイヤー $\exists p (\in P)$ について, 役職割り当ての候補が存在しない場合 ($R(p, i, CR(i)) = \emptyset$), $CCR(i)$ から $CR(i)$ を削除する. $CCR(i) = CCR(i) - \{CR(i)\}$
4	i 日目における全ての役職割り当ての組み合わせ $\forall CR(i) (\in CCR(i))$ において, 任意の役職 $r (\in RR)$ について, 役職候補として r を含むプレイヤーの人数が, 役職 r の上限人数に一致した場合 ($ \{p, R\} \in CCR(i) \mid r \in R = Num(r)$), 役職候補として r を含むプレイヤー $p (\forall p \text{ s. t. } r \in R(p, i, CR(i)))$ に対して, 役職候補の集合 $R(p, i, CR(i))$ 中の役職を r に確定する. $R(p, i, CR(i)) = \{r\}$
5	村人陣営において特殊能力を持つ役職 (占い師, 霊能者, 狩人) の CO を行ったプレイヤー数のうち, 本物の「占い師」, 「霊能者」, 「狩人」の人数を引いた数 (すなわち, 「人狼」もしくは「狂人」のプレイヤー数), および, 「人狼」または「狂人」であると確定されたプレイヤー数の合計 (異なり数) が, 「人狼」と「狂人」の人数の和に一致する場合, 占い師 CO, 霊能者 CO, 狩人 CO を行ったプレイヤー, および, 「人狼」または「狂人」であると確定されたプレイヤーを除く他の全てのプレイヤーの役職候補から, 「人狼」および「狂人」を除外する.
6	三日目までに占い師 CO をしなかったプレイヤーに対して, 四日目以降, 役職の候補集合から「占い師」を除外する. 三日目までに霊能者 CO をしなかったプレイヤーに対して, 四日目以降, 役職の候補集合から「霊能者」を除外する. i 日目に狩人 CO が行われた場合, 狩人 CO をしなかったプレイヤーに対して, $i+1$ 日目において, 役職の候補集合から「狩人」を除外する. 二人のプレイヤーから共有者 CO が行われた後, 共有者 CO をしなかったプレイヤーに対して, 役職の候補集合から「共有者」を除外する.
7	生存しているプレイヤーのうち, 役職が「人狼」であると確定されたプレイヤーの数が, ゲームの勝敗が決しないための人狼の最大人数 (人間の人数の半数以上の最小の自然数 -1) の場合, 役職が「人狼」であると確定されたプレイヤー以外のプレイヤーの役職の候補集合から「人狼」を除外する.

後に, 役職候補として占い師を持つプレイヤーはリーザ一人のみであり, かつ, 役職占い師の最大人数 $Num(\text{占い師}) = 1$ であることから, 表 5 の規則 4 が適用され, リーザの役職が占い師に確定され, リーザに対する役職候補の集合は, $\{\text{占い師}\}$ となる.

7. 関連研究

本論文に関連して, [片上 15] においては, 人狼ゲームの性質の一つとして, 「プレイヤーは自身の役職を客観的に証明することができない」と述べている. これに対して, 本論文では, 人狼ゲームにおけるルールおよび常識をいくつか設定することにより, 人狼ゲーム・ログ中に観測される役職確定情報を手がかりとして他プレイヤーの役職を推論することが可能であることを示した. また, [稲葉 14] においては, 村人陣営, 人狼陣営の各々において, 他人の意見に対する同調・反駁の確率分布を測定し, 村人陣営・人狼陣営を識別する手がかりとして有効であることを示した. [平田 15] においては, 人狼 BBS ログ中の役職宣言発言を抽出し, 人狼エージェントにおいて襲撃対象・処刑対象を選定する確率モデルのパラメータとして組み込む方法を提案した. この研究では, 人狼エージェントにおける行動選択確率モデルのパラメータ調整を目的としているが, 一方, 本論文では, 各プレイヤー視点での役職推論の過程をダイジェストとして可視化することを目的としている.

8. おわりに

本論文では, 人狼ゲームにおける AI 実現の一環として, 人狼ゲームのルールおよび常識を推論規則として用い, 人狼ログ中の役職宣言発言を手がかりとして役職推論を行う過程を人狼ログ・ダイジェストとして可視化する方式を提案し実装した. 今後の課題として, 人狼ログ中の役職宣言発言を自動抽出することが挙げられる.

謝辞

本研究を行うにあたり, 人狼 BBS のデータ使用を許可していただいた ninjin 氏に感謝いたします.

参考文献

- [平田 15] 平田佑也, 稲葉通将, 高橋健一: 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, 第 29 回人工知能学会全国大会論文集 (2015).
- [稲葉 14] 稲葉通将, 島海不二夫, 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, 第 28 回人工知能学会全国大会論文集 (2014).
- [片上 15] 片上大輔, 島海不二夫, 大澤博隆, 稲葉通将, 篠田孝祐, 松原仁, 小林潤平, 関口隆, 新堀英二, 川嶋稔夫: 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).
- [林 16] 林友超, 馬場瑞穂, 宇津呂武仁: 各プレイヤー視点での役職確定情報に基づく人狼ログ・ダイジェストの作成, 言語処理学会第 22 回年次大会論文集, pp. 83-86 (2016).
- [篠田 14] 篠田孝祐, 島海不二夫, 片上大輔, 大澤博隆, 稲葉通将: 汎用人工知能の標準問題としての人狼ゲーム, 第 28 回人工知能学会全国大会論文集 (2014).
- [島海 14] 島海不二夫, 梶原健吾, 大澤博隆, 稲葉通将, 片上大輔, 篠田孝祐: 人狼知能サーバの構築, ゲームプログラミングワークショップ 2014 論文集, pp. 127-132 (2014).