

人狼ゲームプラットフォーム LiCOS を用いた 欺瞞対話コーパスのためのゲームログの収集

阪本浩太郎^{†1†2} 永山翔滋^{†1} 石下円香^{†2} 渋谷英潔^{†1} 森辰則^{†1} 神門典子^{†2†3}

^{†1} 横浜国立大学 ^{†2} 国立情報学研究所 ^{†3} 総合研究大学院大学
E-mail: {sakamoto,nagayama,shib,mori}@forest.eis.ynu.ac.jp,
{ishioroshi,kando}@nii.ac.jp

1 はじめに

人狼ゲームとは、対話を通して「村人」の中に潜伏した「人狼」を見つけ出す対戦型の多人数ゲームであり、近年、人狼知能プロジェクト¹など研究テーマとしても注目されている。「人狼」となったプレイヤーは、他のプレイヤーに正体を悟られないよう、普通の「村人」のふりをして「偽の推理」を披露したり、「占い師（人狼の正体を知ることができる）」を騙って他のプレイヤーを扇動したりすることで勝利を目指す。また、「村人」であっても「人狼」を焙り出すために、敢えて「村人」と認識しているプレイヤーを「人狼」と疑うような発言をする場合もある。

こういった認識している事柄（以降、本心と呼ぶ）と異なる認識を他者に与えようとする発言を本研究では欺瞞と定義する。我々は、これからの対話システムには、ユーザーの命令に唯々諾々と従うのではなく、自発的に対話を進めていくことも必要であると考えており、システムが欺瞞的な発言をすることはその一つに該当すると考えている。こういった研究を進める上で、欺瞞と本心に対応付けた対話コーパス（欺瞞対話コーパス）が必要である。それゆえ、我々は、人狼ゲームにおける対話を収集することで欺瞞対話コーパスを構築しようとしており、そのための人狼ゲームプラットフォーム LiCOS を開発した [1]。

LiCOS のプレイ画面を図 1 に示す。左側に発言のタイムライン、右側にプレイヤーが演じるエージェントと正体の対応表があり、左側の発言（欺瞞）をした時点での本心が右側の対応表に記録できる点が LiCOS の大きな特徴である。本稿では、LiCOS の仕様を説明した後、欺瞞と本心の関係がどのように実際のゲームログに表れているかを報告する。

2 関連研究

従来研究で構築された対話コーパス [2, 3] では、発言者は基本的に誠実であり、本研究の焦点である、相

手を騙すことを目的とした発言は収録されていない。人狼ゲームに関する従来研究 [4, 5] では、人狼 BBS² のログを利用したものが多く、人狼 BBS はオンラインで行う掲示板型人狼ゲームであるが、図 1 に示したような対応表は存在しないため、プレイヤーの本心は役割や投票などの行動から推測するしかない。一方で、

異老 | %五宿 | M修旅虎兵宿妙 † 羊孫娘面
真偽 | 霊白白 | 灰白灰灰灰灰 † 白白狼狼
偽真 | 霊白白 | 白灰灰灰灰灰 † 白白狼狼

のように、人狼 BBS での発言中に表形式で推理や確定事実を述べるプレイヤーは一定数存在し、対応表に相当する形式で自分の推理などを記述している。対応表を導入した人狼プラットフォームは LiCOS の大きな特徴であり、プレイヤーの本心に関するデータを自然に収集できると考えている。

3 LiCOS の仕様

ユーザーはウェブブラウザを通して人間のプレイヤーとしてゲームに直接参加するか、あるいはロボットをゲームに参加させることができる。

図 1 はウェブブラウザ上で人間のプレイヤーが参加したゲームのプレイ画面の例である。左上に、ゲーム内の日にちとフェーズ、そのフェーズ終了までの残り時間を表示する。右上に、ゲーム中にプレイヤーが演じるエージェントの画像、名前（例では「Yihan」）、正体となる役職（例では「Villager（村人）」）と陣営（例では「Villagers（村人陣営）」）を表示する。左中央にプレイヤーの発言のタイムラインを表示する。下部に発言のための発言入力フォームを表示する。下部左側の公開用発言入力フォームから投稿した発言は、ゲーム参加者全体に共有される。下部右側の非公開用発言入力フォームから投稿した発言は、プレイヤー本人のみが閲覧できる。右中央に対応表を表示する。

プレイヤーは、タイムラインを通して他のプレイヤーと対話しつつ、他プレイヤーの正体の予想を対応表で

¹<http://aiwolf.org/>

²<http://ninjinix.com/>



図 1: プレイ画面

容易にまとめることが可能である。対応表の入力を本心と仮定することで、プレイヤーの発話と本心の対応づけが得られる。次に、処刑先投票などで誰に投票したかによっても本心が現れると考えられる。さらに、非公開用発話入力フォームに入力された発話にも、本心が現れる可能性がある。発話、対応表入力、投票は全てタイムスタンプ付きでログとして保存されるため、これらの情報からタイムスタンプを基に発話と本心の対応づけを得る。

4 通信プロトコル

LiCOS のサーバとウェブブラウザ間やサーバとロボット間の通信は WSS (WebSockets over SSL/TLS) 上で行われる。通信データは、JSON-LD 1.1³形式で記述される。JSON-LD 1.1 形式は、全ての値に IRI と呼ばれる識別子を紐付ける規格であり、通信データ内に記述される @id や @context により、値と IRI の対応を記述する。この形式のスキーマは JSON Schema Draft-07⁴で記述されており、通信データの例とスキーマは <https://werewolf.world/>に公開されている。

ソースコード 2 はプレイヤーが送信した発話データの例である。ソースコード 1 はプレイヤーが送信した

対応表入力データの例である。これらのデータの構造について、共通箇所を表 1、異なる箇所を表 2 と表 3 にそれぞれ示す。

ソースコード 1: プレイヤーが送信する対応表入力データの例

```

1 {
2   "@context" : [ "https://werewolf.world/context/0.2/
3     base.jsonld", "https://werewolf.world/context
4     /0.2/board.jsonld" ],
5   "@id" : "https://werewolf.world/resource/0.2/
6     boardMessage",
7   "village" : {
8     "@context" : "https://werewolf.world/context/0.2/
9     village.jsonld",
10    "@id" : "https://licos.online/state/0.2/village",
11    "id" : 8,
12    "name" : "Doubtful Villager",
13    "totalNumberOfAgents" : 5,
14    "lang" : "ja"
15  },
16  "token" : "031387c7-2e91-445f-b6b1-0804
17    b429ebb2",
18  "phase" : "morning",
19  "date" : 2,
20  "phaseStartTime" : "2019-01-14T21:59:03+09:00",
21  "serverTimestamp" : "2019-01-14T21
22    :59:07.052+09:00",
23  "clientTimestamp" : "2019-01-14T12:59:18.195Z",
24  "directionality" : "client to server",
25  "phaseTimeLimit" : 180,
26  "intensionalDisclosureRange" : "private",
27  "extensionalDisclosureRange" : [ ],
28  "myAgent" : {
29    "@context" : "https://werewolf.world/context/0.2/
30    agent.jsonld",
31    "@id" : "https://licos.online/state/0.2/village#8/agent
32    #0",
33    "id" : 0,
34    "name" : {
35      "en" : "Yihan",
36      "ja" : "イーハン"
37    },
38  },

```

³<https://json-ld.org/spec/latest/json-ld/>

⁴<https://json-schema.org/draft-07/json-schema-release-notes.html>

表 1: データの共通する構造

キー	値の説明
village	村の設定
village.id	村の ID
village.name	村の名前
village.totalNumberOfAgents	村の人数
village.lang	村の使用言語
token	プレイヤー識別トークン
phase	現在のフェーズ
date	何日目
phaseTimeLimit	フェーズの時間制限
serverTimestamp	サーバのタイムスタンプ
clientTimestamp	クライアントのタイムスタンプ
directionality	サーバ・クライアント間のデータ送信方向
intensionalDisclosureRange	ゲーム中での公開範囲の内包
extensionalDisclosureRange	ゲーム中での公開範囲の外延 (ログにする際に値が埋まる)
myAgent	送信エージェントの情報
myAgent.id	送信エージェントの ID
myAgent.name	送信エージェントの名前
myAgent.name.en	送信エージェントの英名
myAgent.name.ja	送信エージェントの和名
myAgent.image	送信エージェントの画像
myAgent.role	送信エージェントの役職
myAgent.role.name	送信エージェントの役職名
myAgent.role.image	送信エージェントの役職の画像

表 2: 対応表入力データの構造

キー	値の説明
agent	エージェント
agent.id	エージェント ID
agent.name	エージェントの名前
agent.name.en	エージェントの英名
agent.name.ja	エージェントの和名
agent.image	エージェントの画像
role	役職
role.name	役職の名前
role.name.en	役職の英名
role.name.ja	役職の和名
role.image	役職の画像
prediction	予想 ('?', 'Δ', 'O', 'X')

表 3: 発話データの構造

キー	値の説明
agent	発話エージェント
agent.id	発話エージェントの ID
agent.name	発話エージェントの名前
agent.name.en	発話エージェントの英名
agent.name.ja	発話エージェントの和名
agent.image	発話エージェントの画像
isMine	発話エージェントのプレイヤー識別トークンは token の値と同じか
text	発話テキスト
text.@value	発話テキストの値
text.@language	発話テキストの言語
characterLimit	発話テキストの文字数上限
isOver	このフェーズの発話を終了するか

```

30 "image": "https://werewolf.world/image/0.2/yihan.
    jpg",
31 "role": {
32   "@context": "https://werewolf.world/context/0.2/
    role.jsonld",
33   "@id": "https://licos.online/state/0.2/village#8/role
    #villager",
34   "name": {
35     "en": "Villager",
36     "ja": "村人"
37   },
38   "image": "https://werewolf.world/image/0.2/
    villager.jpg"
39 }
40 },
41 "agent": {
42   "@context": "https://werewolf.world/context/0.2/
    agent.jsonld",
43   "@id": "https://licos.online/state/0.2/village#8/agent
    #4",
44   "id": 4,
45   "name": {
46     "en": "Susan",
47     "ja": "スーザン"
48   },
49   "image": "https://werewolf.world/image/0.2/susan.
    jpg"
50 },
51 "role": {
52   "@context": "https://werewolf.world/context/0.2/
    role.jsonld",
53   "@id": "https://licos.online/state/0.2/village#3/
    role",
54   "name": {
55     "en": "Werewolf",
56     "ja": "人狼"
57   },
58   "image": "https://werewolf.world/image/0.2/
    werewolf.jpg"
59 },
60 "prediction": "X"
61 }

```

ソースコード 2: プレイヤーが送信する発話データの例

```

1 {
2   "@context": [ "https://werewolf.world/context/0.2/
    base.jsonld", "https://werewolf.world/context
    /0.2/chat.jsonld" ],
3   "@id": "https://werewolf.world/resource/0.2/
    playerMessage",
4   "village": {
5     "@context": "https://werewolf.world/context/0.2/
    village.jsonld",
6     "@id": "https://licos.online/state/0.2/village",

```

```

7   "id" : 8,
8   "name" : "Doubtful Village",
9   "totalNumberOfAgents" : 5,
10  "lang" : "ja"
11 },
12 "token" : "031387c7-2e91-445f-b6b1-0804
    b429ebb2",
13 "phase" : "morning",
14 "date" : 2,
15 "phaseStartTime" : "2019-01-14T21:59:03+09:00",
16 "serverTimestamp" : "2019-01-14T21
    :59:37.086+09:00",
17 "clientTimestamp" : "2019-01-14T12:59:39.162Z",
18 "directionality" : "client to server",
19 "phaseTimeLimit" : 180,
20 "intensionalDisclosureRange" : "public",
21 "extensionalDisclosureRange" : [],
22 "myAgent" : {
23   "@context" : "https://werewolf.world/context/0.2/
    agent.jsonld",
24   "@id" : "https://licos.online/state/0.2/village#8/agent
    #0",
25   "id" : 0,
26   "name" : {
27     "en" : "Yihan",
28     "ja" : "イーハン"
29   },
30   "image" : "https://werewolf.world/image/0.2/yihan.
    jpg",
31   "role" : {
32     "@context" : "https://werewolf.world/context/0.2/
    role.jsonld",
33     "@id" : "https://licos.online/state/0.2/village#8/role
    #villager",
34     "name" : {
35       "en" : "Villager",
36       "ja" : "村人"
37     },
38     "image" : "https://werewolf.world/image/0.2/
    villager.jpg"
39   }
40 },
41 "agent" : {
42   "@context" : "https://werewolf.world/context/0.2/
    agent.jsonld",
43   "@id" : "https://licos.online/state/0.2/village#8/agent
    #0",
44   "id" : 0,
45   "name" : {
46     "en" : "Yihan",
47     "ja" : "イーハン"
48   },
49   "image" : "https://werewolf.world/image/0.2/yihan.
    jpg"
50 },
51 "isOver" : false,
52 "characterLimit" : 140,
53 "isMine" : true,
54 "text" : {
55   "@value" : "Susan 怪しくない?",
56   "@language" : "ja"
57 }
58 }

```

5 ゲームログ

LiCOS を用いて 5 人で人狼ゲームをオンラインで行ったゲームログの一部をソースコード 1 とソースコード 2 に示す。2 つのゲームログの `serverTimestamp` の値を比較することで、ソースコード 1 はソースコード 2 より先に入力されていることがわかる。ソースコード 1 は、対応表の「Susan」と「人狼」の対応箇所を「Yihan」が X マークにしたことを示し、「Susan」が「人狼」ではないと「Yihan」が予想したと考えられる。そして、ソースコード 2 は、「Yihan」が公開用発話入力フォームに「Susan 怪しくない?」と入力し、プレイヤー全体に発話を共有したことを示す。従って、

「Yihan」は、対応表の入力が本心であると仮定すれば、本心では「Susan」は「人狼」ではないと考えている。しかし、その後の発話では「Susan」が「人狼」であることを疑っているように表現している。このように、ゲームログから、「Yihan」の欺瞞と本心の関係を見ることが出来る。

6 まとめ

人狼ゲームプラットフォーム LiCOS を開発し、人間同士で人狼ゲームのオンラインプレイがウェブブラウザを通してできるようになった。また、実際のゲームログに現れた欺瞞と本心の関係について、一例を示し、報告した。今後は、欺瞞対話コーパスの作成と公開に向けて、参加者を募り、ゲームログ収集を行いたい。そして、人間だけでなくロボットもゲームに参加できるようにする予定である。

謝辞

この研究は平成 30 年度国立情報学研究所公募型共同研究の助成を受けています。また、本研究を進めるにあたり、サーバマシンを提供してくださった rakumo 株式会社様、開発に協力してくださった横浜国立大学のサークル YNU WAIWAI の宇田川悠大氏に感謝します。

参考文献

- [1] 阪本浩太郎, 渋谷英潔, 森辰則. 欺瞞対話コーパスの構築に向けた人狼ゲームプラットフォーム LiCOS の開発. 言語処理学会第 24 回年次大会発表論文集, pp. 885-888, 2018.
- [2] 東中竜一郎, 船越孝太郎, 荒木雅弘, 塚原裕史, 小林優佳, 水上雅博. テキストチャットを用いた雑談対話コーパスの構築と対話破綻の分析. 自然言語処理, 23(1), 2015.
- [3] 塚原裕史, 内海慶. オープンプラットフォームとクラウドソーシングを活用した対話コーパス構築方法. 言語処理学会第 21 回年次大会発表論文集, 2015.
- [4] 林友超, 馬場瑞穂, 宇津呂武仁. 役職確定情報に着目した人狼ログ・ダイジェストの作成. 第 30 回人工知能学会全国大会論文集, 2016.
- [5] 稲葉通将, 鳥海不二夫, 高橋健一. 人狼ゲームデータの統計的分析. ゲームプログラミングワークショップ 2012 論文集, 2012.