人狼エージェントにおける深層Qネットワークの応用

王 天鶴[1,a)]　金子 知適[2,3,b)]

概要：人狼ゲームとは，不完全情報ゲームの一種である．人狼ゲームをプレイするエージェントは，主に2つの問題に対応する必要がある．本研究では，一つは「投票や特殊能力の対象選択」の問題であり，もう一つは「情報交換」の問題である．本論文は前者の問題に注目する．一つ目の問題に於いて，エージェントは「どのプレイヤに投票するか」,「どのプレイヤに特殊能力を使うか」を決める．我々は人狼エージェントに深層Qネットワークの技術を応用する．それに，Q学習を応用した既存エージェントに基づいて新しい状態表現と行動表現を提案する．提案したエージェントは既存のQ学習エージェント，人狼知能大会に参加したエージェントと性能を比較して評価する．
実験では393個ニューロン（執筆時点まで人狼ゲームエージェントに応用したQネットワークの最大ニューロン数である）のQネットワークを持つエージェントを50万回のゲームで学習させた結果を評価した．実験結果によると，同じ対戦環境において，提案したエージェントは，Q学習を応用したエージェントと一部の人狼知能大会参加プログラムより勝率が高いと評価された．「投票や特殊能力の対象選択問題」に対して，ヒューリスティックスな手法を使わず，学習手法のみを利用して有意義な結果を得られた．

Application of Deep Q Network in Werewolf Game Agents

Tianhe Wang[1,a)]　Tomoyuki Kaneko[2,3,b)]

Abstract: Werewolf, also known as Mafia, is a kind of game with imperfect information. Werewolf game agents must cope with two kinds of problems, “decision on who to trust or to kill”, and “decision on information exchange”. In this paper, we focus on the first problem. In the first problem, players decide to select a target to trust or to kill. We consider werewolf game as a Markov decision process and propose a method to use techniques in deep Q network to build werewolf agents. We proposed new representation of states and actions based on existing agents trained by Q learning method. Our proposed agents were compared with existing agents trained by Q learning method and with existing agents submitted to the AIWolf Contest, the most famous werewolf game agents contest in Japan.
In our experiment, we evaluate our agent with Q network of 393 neurons (Q network with most neurons in werewolf agents until we write this paper) after learning for 500000 games. Experimental results showed that, when agents learned and played with same group of players, our proposed agents have better player performances than existing agents trained by Q learning method and a part of agents submitted to the AIWolf Contest. We obtained promising results by using reinforcement learning method to solve “decision on who to trust or to kill” problem without using heuristic methods.

[1] 東京大学大学院学際情報学府
Graduate School of Interdisciplinary Information Studies, The University of Tokyo
[2] 東京大学大学院情報学環
Interfaculty Initiative in Information Studies, The University of Tokyo
[3] 国立研究開発法人科学技術振興機構さきがけ
JST, PRESTO
[a)] wangtianhe@g.ecc.u-tokyo.ac.jp
[b)] kaneko@acm.org

1. はじめに

近年，人工知能は囲碁等のような完全情報ゲームで良い成績が報告されている [4]．不完全情報ゲームの人工知能についての研究はより良い成果が期待されている．

本研究の目的は不完全情報ゲームの人狼ゲームを題材とし，人間の知識をなるべく使わずに人狼エージェントの強さを検証することである．そのために深層Qネットワー

ク [3] の技術を応用する．人狼知能大会 [1] に出場したエージェント，出場せずに論文で公表された従来手法と比較することを性能評価として行う．

2. 研究背景

2.1 人狼ゲーム

人が遊ぶための人狼ゲームは自然言語で進行するゲームであるが，コンピュータエージェントが競うために，進行をプロトコルで定めた対戦プラットフォーム [7] も存在する．

人狼ゲームにおいて，村人，占い師，狂人，人狼のような様々な役職が存在する．各役職は人狼陣営と人間陣営という二つの陣営に所属する．例えば，村人と占い師は人間陣営に所属し，人狼と狂人は人狼陣営に所属する．もし全ての村人が排除されたら，人狼陣営は勝利する．逆に，もし全ての人狼が排除されたら，人間陣営は勝利する．全てのプレイヤは他のプレイヤに投票することができる．デフォルト設定の 5 人ゲームは 2 人の村人，1 人の占い師，1 人の狂人，1 人の人狼で構成される．どのプレイヤがどの役職であるかは，ゲーム終了まで公開されない．

人狼ゲームは，昼と夜という 2 つの段階に分けることができる．夜に，占い師と人狼が他のプレイヤに特殊能力を使うことができる．昼に，各プレイヤは会話で情報を交換することで相手を説得したり自身の疑惑を晴らせるよう努力する．例えば，プレイヤは自分の役職を公表することが出来る．これはカミングアウト（以下は CO とする）と呼ぶ．しかし，嘘をつくことも許されるので，他のプレイヤは CO を信頼するかどうか判断する必要がある．会話内容はプロトコルで定義した言葉で構成される．会話後，各プレイヤは自己意志で投票し，最大票数のプレイヤは排除される．

人狼ゲームのエージェントが賢く振る舞うためには，2 種類の問題に対応する必要がある．1 つ目は「どのプレイヤに投票するか」,「どのプレイヤに特殊能力を使うか」を決めることであり，本稿では,「投票や特殊能力の対象選択問題」と呼ぶ．2 つ目は会話する時にプレイヤは「どのような情報を交換することで相手を説得したり自身の疑惑を晴らせるか」を決めることであり，本稿では「情報交換問題」と呼ぶ．本研究では，1 つ目の問題を注目する．

2.2 強化学習

強化学習 [5] は，エージェントが環境における一番高い報酬値を得る方策を学習する手法である．エージェントは，与えられた状態に対する行動を選択し，行動を実行した後の環境から次の状態と報酬を受け取ることを繰り返す．この環境から受け取る報酬の累積和を最大化できるように学習を行うことが強化学習である．

Q 学習は強化学習手法の一つである．Q 値はある状態においてある可能な行動をとった時の，今後得られる報酬を推定する値である．各状態における行動の Q 値を表で格納し，エージェントの試行錯誤を通じて Q 値を反復的に更新することで学習する．

2.3 深層 Q ネットワーク

深層 Q ネットワーク [3] は，表の代わりに深層ネットワークで Q 値を近似するモデルである．環境から観測する情報が多く状態空間がかなり大きいゲームの場合，表で Q 値を格納することは現実ではない．表の代わりに深層ネットワークを利用して Q 学習の動作が可能であり，性能も向上できる．

オンラインの Q 学習と異なり，性能向上のために経験再生という機制が導入される．経験再生は過去の経験からランダムでサンプリングし，経験を繰り返し利用して訓練することである．

深層 Q ネットワークは既に Atari 等のゲームのエージェントで応用されたことがあり，良い成績が報告されている．

3. 関連研究

多くの人狼ゲームエージェント研究者 [8] [6] は人狼知能大会が公開したログから，プレイヤの役職の推定能力を教師あり学習手法で学習し，その推定結果に基づいてヒューリスティックな策略で行動を取る．

強化学習を人狼ゲームに応用した研究は 2 件存在する．

研究 [7] ではプラットフォーム [9] において人狼ゲームエージェントでの Q 学習手法を提案した．研究 [7] のエージェントは，状態表現を「会話内容に基づいて推測した全てのプレイヤの可能な役職の組み合わせ」とし，行動を「能力者 CO する条件の選択」,「人狼，狂人の騙り役職の選択」,「投票，占い，護衛，攻撃の対象選択」という 3 項目で扱った．彼らのエージェントは下記の問題を解決するのを目標とする：

- 「情報交換問題」の子問題
  - どんな条件が満足できたら CO するか？
  - 自分が人狼，狂人の場合はどの役職として CO するか？
- 「投票や特殊能力の対象選択問題」の子問題
  - どのプレイヤに投票するか？
  - どのプレイヤに特殊能力を使うか？

条件と選択のルールはヒューリスティックで定義された．

本研究は，ヒューリスティックを用いずに，行動選択で全行動を対象とする．これにより利用可能な情報の大部分が利用できて，エージェントの性能が制限されないという期待ができる．

同じく Q 学習を応用した研究 [10] は,「投票や特殊能力の対象選択問題」のみを解決するのを目標とする．

彼らは状態表現を表 1 と 2 が示すように，会話から抽出

表 1　研究 [10] の状態表現

	プレイヤ 1	プレイヤ 2	プレイヤ 3
プレイヤ 1	宣言 (1)	事実 (2)	事実 (3)
プレイヤ 2	態度 (2,1)	宣言 (2)	態度 (2,3)
プレイヤ 3	態度 (3,1)	態度 (3,2)	宣言 (3)

表 2　研究 [10] の状態表現に含まれる特徴

特徴	説明	値
宣言 (n)	プレイヤ n の CO 役職である.	人間陣営 / 人狼陣営 / 無し
態度 (m, n)	プレイヤ m がプレイヤ n について持っている態度である. 態度は会話内容で判断する.	信用する / 怪しい / 無し
事実 (n)	プレイヤ n に関する事実である. プレイヤ n の役職に指す.	役職名 / 無し

した態度，宣言，事実という 3 項目とする.

行動の選択肢をヒューリスティックに新しく定義された対象選択の策略とした.

- 一番疑っているプレイヤを選ぶ.
- 一番疑われるプレイヤを選ぶ.
- 反対態度を持つ 2 人のプレイヤの中から 1 人を選ぶ.
- 発言が一番多いプレイヤを選ぶ.
- 発言が一番少ないプレイヤを選ぶ.
- ランダムにプレイヤを選ぶ.

投票と特殊能力を 2 種類の行動としてそれぞれ学習させる.

状態表現は局面内容をある程度豊かに表現できるようになるが，ヒューリスティックに定義された行動表現は研究 [7] と同じ問題点が存在する.

本稿の執筆まで，深層 Q ネットワークを人狼ゲームに応用した研究の例は検索できない．本研究は初回として人狼ゲームに深層 Q ネットワークを適用する可能性を検討する.

4. 提案手法

本研究では，ヒューリスティックな手法を使わず，強化学習で「投票や特殊能力の対象選択問題」を解決することを目指す．研究 [10] を拡張して，状態表現と行動表現を改善し，深層 Q ネットワークの技術で学習するエージェントを提案する.

4.1 Q ネットワークの構造

本研究で用いる状態表現には，既存研究より豊かな情報が反映されているので，状態数も大幅に増加すると想定される．状態数の多い状況にも Q 学習の性能を向上するために，深層 Q ネットワークの技術で学習するのを提案する.

Atari のゲームの入力は画像なので，DQN では畳み込み層を用いて画像を処理したが，人狼ゲームの入力は画像ではないので，畳み込み層を用いることは適切でない．図 1 の示すように，畳み込み層の代わりに，全結合層で Q ネットワークを構成した.

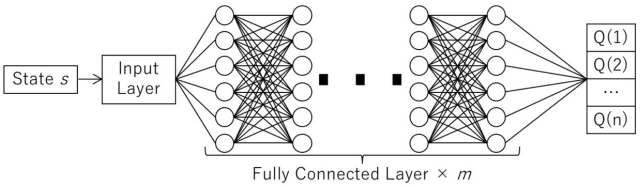

図 1　提案した Q ネットワーク

表 3　態度にある特徴

特徴	説明	値
特殊能力 (m, n)	プレイヤ m がプレイヤ n に特殊能力を使うと言ったかどうか.	言った / 言っていない
投票 (m, n)	プレイヤ m がプレイヤ n に投票すると言ったかどうか.	言った / 言っていない
疑う (m, n)	プレイヤ m がプレイヤ n の身分を疑うかどうか.	疑う / 疑わない
投票済み (m, n)	プレイヤ m がプレイヤ n に投票したと言ったかどうか.	言った / 言っていない

村人と狂人の場合は，各エージェントが 1 個の Q ネットワークを持ち，投票する際に利用する．人狼と占い師の場合は，各エージェントが 2 個の Q ネットワークを持ち，それぞれを投票する際と，特殊能力を使う際に利用する.

4.2 状態表現

適切で有用な状態表現を設計することは，深層 Q ネットワークの技術を人狼ゲームに適用するための課題の一つである.

エージェントがゲームの各局面に対応した適切な行動を取るためには、エージェントが用いる状態表現に局面の情報が適切に反映されている必要がある。本手法では，状態表現において，できるだけ多い情報を入れる.

提案する状態表現において，3 種類の情報が含まれる．態度は表 3 の示すように 4 つの特徴とする．事実は表 4 の示すように 3 つの特徴とする．宣言は表 5 の示すように 3 つの特徴とする．全ての特徴値が整数でコーディングされる.

拡張された 3 種類の情報は表 1 のように行列を構成する上で，行列をベクトルに変形する．ゲーム進行の日付をベクトルに追加し，本研究が提案する状態表現になる.

4.3 行動表現

「投票や特殊能力の対象選択問題」において，行動は目標プレイヤを決定することである.

行動は「プレイヤ x を選ぶ」にする．その理由は，行動が選択肢を全てのプレイヤをカバーすると，エージェント

表 4　事実にある特徴

特徴	説明	値
役職 (n)	プレイヤ n の役職である．（エージェントが占い師の場合，ここはプレイヤ n の占い結果に指す）	役職名 / 無し
投票 (n)	エージェントはプレイヤ n に投票したかどうか.	した / していない
特殊能力 (n)	エージェントはプレイヤ n に特殊能力を使ったかどうか.	使った / 使っていない

表 5　宣言にある特徴

特徴	説明	値
CO(n)	プレイヤ n が CO した役職である.	無し / 役職名
生存 (n)	プレイヤ n が生存しているかどうか.	生存 / 死亡
自己 (n)	プレイヤ n がエージェントであるかどうか.	である / でない

の性能を制限する程度が低くなると期待される．

n 人ゲームにおいて，エージェントは最大，$n-1$ 個の行動選択肢を選べる．つまり，エージェント自身以外のプレイヤは全部目標プレイヤとして選べる．

5. 実験と評価

この部分では，提案手法で学習させたエージェントを，他のプレイヤに対する勝率で評価する．

学習段階と評価段階という 2 つの段階で実験を構成する．学習段階ではエージェントの学習の速度を観測し，評価段階では提案手法を用いるエージェントを他のエージェントと比較する．

5.1 エージェント構成

エージェントは [2] の Python 用フレームワークを用いて構築した．

ゲームにおいて，投票する際と特殊能力を使う際に，エージェントが状態を観測する．観測した状態は 4.1 節で提案した Q ネットワークに入力され，最終的に各目標プレイヤを選択する Q 値が計算される．学習段階では，エージェントは ϵ-グリーディ法を利用して行動を取る．ϵ は学習が始めてからあるエピソードまで初期値から最終値までエピソード数に比例して下がる．評価段階では，ネットワークのパラメータの更新を止め，グリーディ法で行動を選択する．

報酬値はゲームの状況に基づいて与える．ゲームがエージェント側の勝利で終了した場合に，報酬値が 10 と与えられる．それ以外の場合，すなわちゲーム継続中の場合やエージェント側の敗北で終了した場合は報酬値が 0 である．

また，エージェントはシンプルな策略で情報交換をする．占い師の場合に，エージェントは最初から自動的に CO し，占い結果を公表する．村人の場合に，他のプレイヤに疑われたら村人として CO する．狂人の場合に，他のプレイヤに疑われたら占い師として CO するが，占い結果に関する発言はない．人狼の場合に，他のプレイヤに疑われたら村人として CO する．

5.2 実験設定

ゲーム人数は 5 人である．全てのプレイヤが AIWolf Server [9] に繋がって人狼ゲームを始める．学習段階において，1 人のプレイヤは提案のエージェントであり，他のプレイヤは全員 AIWolf Server が提供した Sample Player とした．

5 人人狼ゲームにおいて，プレイヤは 2 人の村人，1 人の占い師，1 人の人狼と 1 人の狂人である．この実験では，4 つのエージェントを 4 つの役職としてそれぞれ学習させる．

実験で利用する Q ネットワークは入力層，4 層の全結合層と出力層で構成される．入力層は 101 個のニューロンを持ち，4 層の全結合層はそれぞれ 128 個，64 個，64 個，32 個のニューロンを持ち，出力層は 4 個のニューロンを持つ．全ての全結合層は ReLU を活性化関数とする．

各エージェントは 50 万回のゲームで学習する．ϵ は 40 万目のタイムステップまで下がる．初期値は 0.9 とし，最終値は 0.2 とした．ネットワークの重みは 8 つのタイムステップ毎に更新される．学習率は 0.001 とした．経験再生のメモリは 45 万タイムステップとした．Q 学習の γ は 0.99 とした．

評価段階において，比較するため，学習なしのエージェント，Sample Player，研究 [10] のエージェント，人狼知能大会参加者のエージェントの性能も観測する．表 6 の示すように，近年 3 回の人狼知能大会（GAT2017，CEDEC2017，GAT2018）において，Java 言語で実装された参加者エージェントからそれぞれ上位の 3 人を比較対象とする．

各エージェントの性能は下記の 2 つの対戦環境においての勝率で評価される．各プレイヤの役職は一様ランダムで割り当てられる．

対戦環境 I : 5 人のプレイヤに，1 人を評価対象のエージェントとし，他の 4 人を全員 Sample Player とする．

対戦環境 II : 5 人のプレイヤに，1 人を観測するエージェントとし，他の 4 人を 1 人の 2017gat-1st，1 人の 2017cedec-1st，1 人の 2017cedec-2nd と 1 人の sample player とする．

5.3 実験結果

学習段階において，提案のエージェントと研究 [10] のエージェントの勝率を 1000 回ゲーム毎に測定した．勝率の変化は図 2，3，4，5 に示す．図において，評価対象が 1000 回のゲーム毎の勝率値で描画され，比較対象が水平線

表 6　比較対象とした参加者エージェント

順位	人狼知能大会		
	GAT2017	*CEDEC2017*	*GAT2018*
1 位	m_cre	cndl	rsaito
2 位	wasabi	kasuka	neko
3 位	tori (Sample Player)	wasabi	wasabi

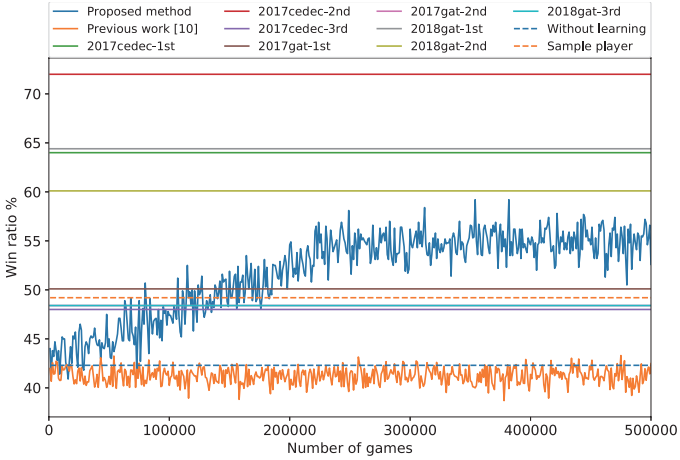

図 2　狂人エージェントの勝率変化

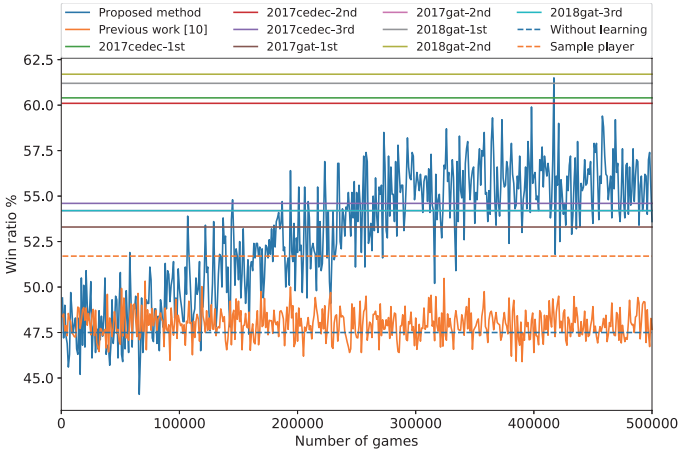

図 3　占い師エージェントの勝率変化

で描画されている．

全ての役職に対して，学習が始める際に，提案のエージェントは研究 [10] の勝率はほぼ同様であることが分かる．

狂人の場合，提案エージェントの勝率は凡そ 12 ポイント上昇した．最終の勝率は 2017cedec-1st，2017cedec-2nd，2018gat-1st，2018gat-2nd 以外のエージェントの勝率を超えた．研究 [10] のエージェントの勝率は全体的に学習なしのエージェントとほぼ同じである．

占い師の場合，提案エージェントの勝率は凡そ 8 ポイント上昇した．最終の勝率は 2017cedec-1st，2017cedec-2nd，2018gat-1st，2018gat-2nd 以外のエージェントの勝率を超えた．研究 [10] のエージェントの勝率は全体的に学習なしのエージェントとほぼ同じである．

村人の場合，提案エージェントの勝率は凡そ 14 ポイント上昇した．最終の勝率は 2017cedec-2nd 以外のエージェントの勝率を超えた．研究 [10] のエージェントの勝率は全体的に学習なしのエージェントとほぼ同じである．

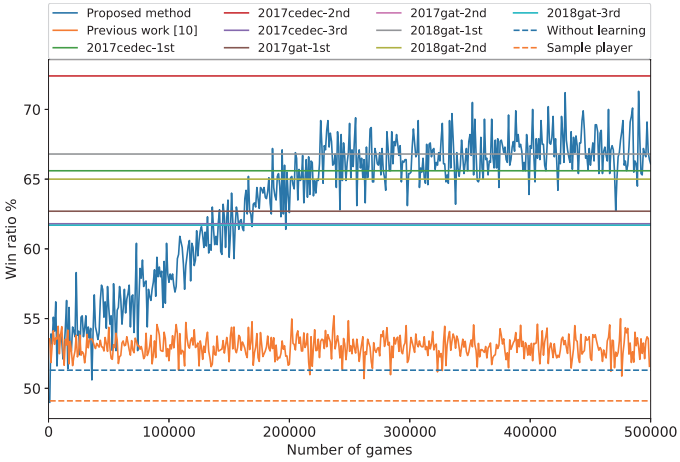

図 4　村人エージェントの勝率変化

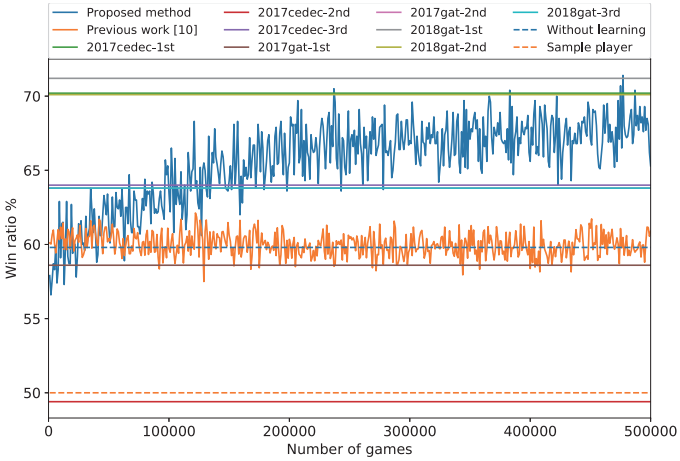

図 5　人狼エージェントの勝率変化

人狼の場合，提案エージェントの勝率は凡そ 10 ポイント上昇した．最終の勝率は 2017cedec-1st，2018gat-1st，2018gat-2nd 以外のエージェントの勝率を超えた．研究 [10] のエージェントの勝率は全体的に学習なしのエージェントとほぼ同じである．

学習段階において，提案のエージェントは研究 [10] より性能が良いと評価できた．

評価段階の結果は表 7，8，9，10 のように示す．

全ての役職において，提案のエージェントは対戦環境 I においての勝率が対戦環境 II においての勝率より高いと観測した．

狂人，占い師，村人の場合，対戦環境 II においての順位は対戦環境 II より大幅に落ちたと観測した．特に占い師の順位は 5 位から 12 位に落ちた．

人狼の場合，対戦環境 II においての勝率が対戦環境 I においての勝率より下がったが，順位は 2 位のままで変わらなかった．

また，2 つの対戦環境において，研究 [10] のエージェン

表 7　狂人エージェントの勝率

評価対象の	対戦環境 I		対戦環境 II	
エージェント	勝率	順位	勝率	順位
提案手法	56.5%	5	26.2%	8
研究 [10]	43.0%	10	15.1%	11
GAT2017-1st	50.1%	6	26.2%	9
GAT2017-2nd	39.9%	12	14.3%	12
CEDEC2017-1st	64.0%	3	49.0%	1
CEDEC2017-2nd	72.0%	1	42.5%	4
CEDEC2017-3rd	48.0%	9	27.9%	7
GAT2018-1st	64.4%	2	48.9%	2
GAT2018-2nd	60.1%	4	48.4%	3
GAT2018-3rd	48.4%	8	28.3%	6
Sample player	49.2%	7	28.9%	5
学習なし	42.3%	11	19.3%	10

表 8　占い師エージェントの勝率

評価対象の	対戦環境 I		対戦環境 II	
エージェント	勝率	順位	勝率	順位
提案手法	57.4%	5	52.9%	12
研究 [10]	48.7%	11	56.8%	8
GAT2017-1st	53.3%	9	54.4%	11
GAT2017-2nd	55.6%	6	63.1%	3
CEDEC2017-1st	60.4%	3	61.1%	6
CEDEC2017-2nd	60.1%	4	60.1%	7
CEDEC2017-3rd	54.6%	8	63.5%	2
GAT2018-1st	61.2%	2	62.4%	4
GAT2018-2nd	61.7%	1	61.6%	5
GAT2018-3rd	54.2%	7	65.0%	1
Sample player	51.7%	10	56.2%	9
学習手法	47.5%	12	55.4%	10

表 9　村人エージェントの勝率

評価対象の	対戦環境 I		対戦環境 II	
エージェント	勝率	順位	勝率	順位
提案手法	71.9%	2	56.1%	8
研究 [10]	53.0%	10	54.6%	10
GAT2017-1st	62.7%	7	53.6%	11
GAT2017-2nd	65.5%	5	64.8%	6
CEDEC2017-1st	65.6%	4	72.4%	2
CEDEC2017-2nd	72.4%	1	72.5%	1
CEDEC2017-3rd	61.8%	8	64.3%	7
GAT2018-1st	66.8%	3	71.6%	4
GAT2018-2nd	65.0%	6	72.2%	3
GAT2018-3rd	61.7%	9	66.1%	5
Sample player	49.1%	12	53.4%	12
学習なし	51.3%	11	55.7%	9

トの順位は 8 位から 1 1 位までである．人狼以外の場合，対戦環境 II においては提案のエージェントの勝率と近いと観測した．人狼の場合に，提案のエージェントの勝率は 2 つの対戦環境において研究 [10] のエージェントより高い．

5.4　結果分析

実験結果によると，提案のエージェントは研究 [10] のエージェントより性能が良いと観測した．原因を明らかにするために，エージェントが学習段階において経験した状態数を観測した．

表 10　人狼エージェントの勝率

評価対象の	対戦環境 I		対戦環境 II	
エージェント	勝率	順位	勝率	順位
提案手法	71.1%	2	52.0%	2
研究 [10]	57.8%	10	42.6%	8
GAT2017-1st	58.6%	9	47.6%	7
GAT2017-2nd	61.2%	7	42.0%	9
CEDEC2017-1st	70.2%	3	53.7%	1
CEDEC2017-2nd	49.4%	12	33.1%	12
CEDEC2017-3rd	64.0%	5	48.1%	6
GAT2018-1st	71.2%	1	51.6%	3
GAT2018-2nd	70.2%	4	49.7%	4
GAT2018-3rd	63.8%	6	49.2%	5
Sample player	50.0%	11	35.5%	11
学習なし	59.8%	8	40.1%	10

図 6 の示すように，提案のエージェントは研究 [10] のエージェントより状態を多く経験できた．状態はエージェントが識別できるゲーム局面を表現すると考える．状態数が増えると，ゲーム局面が細かく正確に識別できるようになる．それがエージェントの決定能力にとってメリットがあると考える．

6.　結論

本研究は，ヒューリスティック手法を使わず，「投票や特殊能力の対象選択問題」を解決するために，深層 Q ネットワークの技術を人狼ゲームエージェントの構築に応用した．提案手法は研究 [10] の手法に基づいて改善した手法である．主に状態表現と行動表現をメインとして改善した．それに，Q ネットワークを用いて Q 値を近似し，経験再生を学習に導入した．

実験結果によると，提案したエージェントは，同じ対戦環境において学習すれば，学習後の勝率は研究 [10] のエージェントと一部の人狼知能大会参加者エージェントより高い．

将来は，エージェントが多様な環境で適応できる能力を強化するように改善する．今回提案したエージェントは学習後に新しい対戦相手と対戦した際に性能が悪くなった．多様な環境でエージェントを学習させること等を計画する．また，エージェントを複数な環境で同時に学習させる非同期な強化学習手法も改善方針として導入できると考える．

謝辞

この研究の一部は，JSPS 科研費 16H02927 と JST さきがけの支援を受けています．

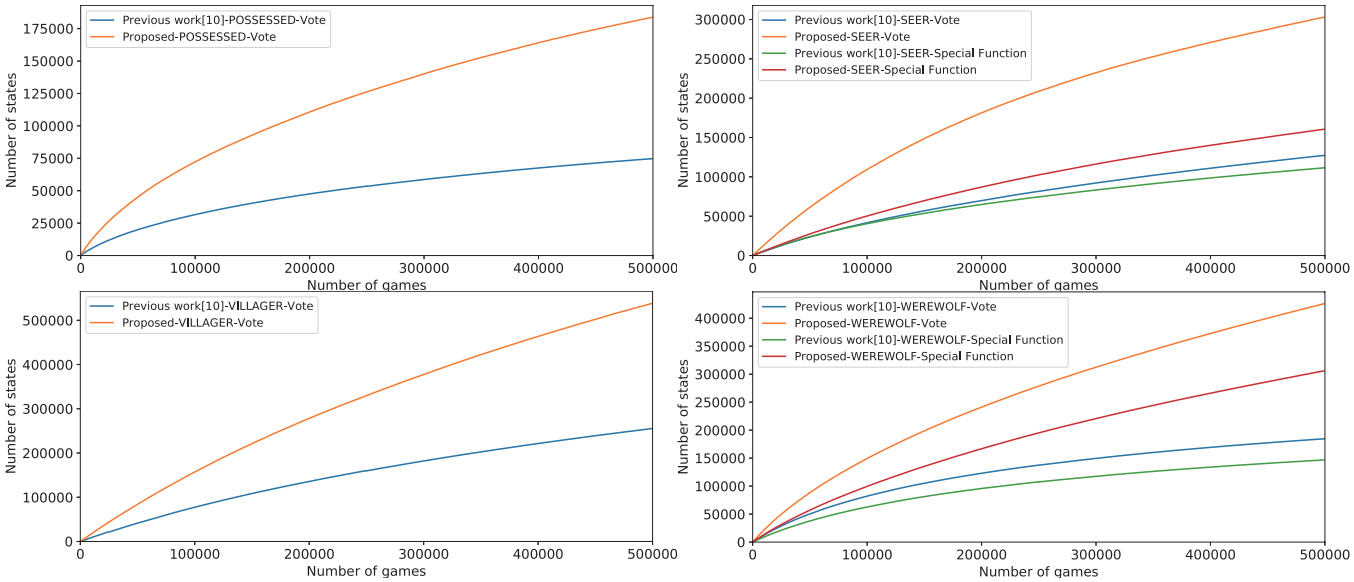

図 6 学習段階において経験した状態数の変化

参考文献

[1] AIWolf: 人狼知能大会 — Artificial Intelligence based Werewolf, `http://aiwolf.org/aiwolf_contest`. Accessed October 10, 2018.

[2] k-harada: k-harada/AIWolfPy, `https://github.com/k-harada/AIWolfPy`. Accessed October 10, 2018.

[3] Mnih, V. et al.: Human-level control through deep reinforcement learning, *Nature*, Vol. 518, No. 7540, p. 529 (2015).

[4] Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M. et al.: Mastering the game of Go with deep neural networks and tree search, *nature*, Vol. 529, No. 7587, p. 484 (2016).

[5] Sutton, R. S. and Barto, A. G.: *Reinforcement learning: An introduction*, Vol. 1, No. 1, MIT press Cambridge (1998).

[6] 大川貴聖，吉仲　亮，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，ゲームプログラミングワークショップ 2017 論文集，Vol. 2017, pp. 50–55 (2017).

[7] 梶原健吾，鳥海不二夫，稲葉通将：人狼における強化学習を用いたエージェントの設計，人工知能学会全国大会論文集，Vol. 29, pp. 1–3 (2015).

[8] 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，狩野芳伸：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，人工知能学会全国大会論文集，Vol. JSAI2016, pp. 2F41–2F41 (2016).

[9] 鳥海不二夫，梶原健吾，大澤博隆，稲葉通将，片上大輔，篠田孝祐：人狼知能プラットフォームの開発，日本デジタルゲーム学会 (2015).

[10] 王　天鶴，金子知適：人狼ゲームエージェントにおける行動選択手法の比較，ゲームプログラミングワークショップ 2017 論文集，Vol. 2017, pp. 177–182 (2017).