The 32nd Annual Conference of the Japanese Society for Artificial Intelligence, 2018

Differentiable Neural Computerを用いた人狼知能の開発

Development of Werewolf-game Agent Using Differentiable Neural Computer

家原　瞭　Ryo Iehara
廣田　敦士　Atsushi Hirota
田中　一晶　Kazuaki Tanaka
荒木　雅弘　Masahiro Araki
岡　夏樹　Natsuki Oka

京都工芸繊維大学
Kyoto Institute of Technology

Werewolf-game is a table game with imperfect information. It is a new challenge for artificial intelligence to play Werewolf-game. We developed a Werewolf-game AI using Differentiable Neural Computer (DNC) (Graves et al., 2016). The system predicted the actions of Werewolf-game agents. As an evaluation index of prediction, we used whether information identifying the behavior matches that of the actual behavior. The evaluation result showed that speech behavior could be predicted with an accuracy of 66.8%. However, behaviors other than speech could be predicted only at the chance level.

1. 緒言

人工知能の分野では昔からチェスや将棋，囲碁といったゲーム AI の研究が盛んにおこなわれている．最近では 2017 年に Google DeepMind の AlphaGo[Silver 16] が柯潔九段に囲碁で勝利したことが話題になった．囲碁という情報空間の広いゲームにおける人工知能の勝利によって，完全情報ゲームでは人間を超えたと言える．そこでゲーム AI の次の挑戦として人狼ゲームが挙げられている [鳥海 16][篠田 14]．鳥海らは，人狼ゲームを人工知能分野の新たなグランドチャレンジとしている．具体的には，AI 同士での人狼ゲームの対戦を行えるようなサーバーの開発 [鳥海 14] や，大会の開催などを行い，様々な方向から人狼ゲームを行う AI（以下人狼知能）の発展を目指している．人狼ゲームは不完全情報ゲームであり，またコミュニケーションを通してやり取りするという点で，今までの人工知能研究の対象と性質が違うゲームといえる．本研究では，人狼知能に対する新たなアプローチとして完全学習型のエージェントの開発を目指す．完全学習型とは，行動の選択や発言内容までエージェントの行動選択を全て学習して選択するものを指す．また，本論文ではその第一段階として現存する強いエージェントの行動を学習のみで予測できることを示す．

2. 人狼ゲーム

2.1 基本ルール

人狼ゲームは 10 数人のプレイヤーによって行われるパーティーゲームである．ゲームスタート時に各プレイヤーには役職が割り振られる．役職ごとに陣営があり，プレイヤーは各陣営に分かれる．陣営ごとの役職は以下のようになっている．

村人陣営： 村人，占い師，霊媒師，狩人

人狼陣営： 人狼，狂人

最終的には以下の勝利条件を満たしたチームが勝利となる．

村人陣営の勝利条件： 人狼をすべて追放する．

人狼陣営の勝利条件： 村人の数を人狼以下にする．

連絡先: 家原　瞭，京都工芸繊維大学，iehara@ii.is.kit.ac.jp

人狼ゲームは，ゲームの世界における半日単位で進行し，以下のような行動が行われる．

昼： 各プレイヤー間で自由に対話を行い，その後全プレイヤーの投票により，人狼容疑者 1 名を追放する．

夜： 人狼により村人が一人襲撃され殺される．役職による能力の使用もこの時に行われる．

追放および襲撃にあったプレイヤーは，それ以降ゲームには参加できない．よって追放と襲撃によってプレイヤーが減っていき，いずれ村人陣営と人狼陣営のどちらかの勝利条件が満たされることになる．各勝利条件を満たすための方針として，村人陣営は昼の対話や占い師や霊媒師の能力を使い，人狼を見つけて，その人物を投票によって追放することを目指す．また，村人陣営のほうが人数が多いため，人狼陣営は自分が人狼であるとばれないように相手をだまし，追放によって村人陣営のプレイヤーを減らしていく必要がある．

2.2 役職説明

村人： 特別な能力なし．

占い師： 夜のフェーズに 1 人のプレイヤーを指定して，人狼か否かを知ることができる．

霊媒師： 夜のフェーズに昼のフェーズに追放したプレイヤーが人狼か否かを知ることができる．

狩人： 夜のフェーズに 1 人のプレイヤーを指定して，護衛することができる．守られているプレイヤーが人狼による襲撃にあった場合，襲撃は失敗となりプレイヤーが減ることはない．

人狼： 夜に 1 人のプレイヤーを指定して襲撃することができる．襲撃されたプレイヤーは以降ゲームに参加できなくなる．また，人狼同士は互いが人狼であることを知っていて，夜のフェーズに人狼内で秘密の会話を行うことができる．

狂人： 特別な能力なし．人狼陣営であるため，人狼にうまく協力する必要がある．

2.3 ゲーム AI としての人狼ゲーム

人狼ゲームはパーティーゲームとして広く受け入れられているが，ゲーム AI の題材としてはどうであろうか．緒言で述べたように，完全情報ゲームでは人間の到達できるレベルを超えてしまったと考える．一方で，ゲームの中には情報の一部が全体に公開されないようなゲームも多く存在する．このようなゲームを不完全情報ゲームと呼ぶ．人狼ゲームはこの不完全情報ゲームであり，新たな挑戦として取り組まれているものの一つである．また，人狼ゲームはコミュニケーションのみによって成り立ち，盤面などの評価が存在しないという点で，他のゲームとは大きく異なる．自然言語対話の中で論理や思考を読み取り，嘘や説得といった要素を含む高度なゲームであり，自然言語対話などに向けた足がかりとなる人工知能のグランドチャレンジとみなされている [鳥海 16][篠田 14]．

2.4 人狼知能にかかわる研究

人狼知能プロジェクト [鳥海] では，人間同士の対戦における行動の解析や人狼知能の大会などを開いている．人間同士の対戦における行動の解析では，熟練度によって勝率が上昇することが示されている．じゃんけんなどの学習不可能な運にのみ依存するゲームでは人工知能にプレイさせることに意味がないが，人狼ゲームは熟練の要素が存在することが示されている [鳥海 16]．

また，同じく人狼知能プロジェクトは，集団知による人狼知能の構築を目指して人狼知能大会を開催している．人狼知能大会は 2017 年に第 3 回人狼知能大会が開催されている．そこでは，人狼ゲーム用の会話プロトコルを用意しており，そのプロトコルを用いた大会が行われている．また，第 3 回からは自然言語を用いた大会も行われている．それにともなって人狼知能エージェントを戦わせることのできるサーバーとそのサーバーで戦わせることのできるサンプルの構築と公開を行っている [鳥海 14]．

3. 構築した学習システム

3.1 システムの目的

本研究で作成したシステムは，人狼ゲームのシステム同士での対戦におけるログデータから，プレイヤーの行動を予測できるよう学習することが目的である．ログデータから一人のプレイヤーを選択し，そのプレイヤー以外の行動のうち観測可能な情報を入力とし，そのプレイヤーの行動を推測する．

3.2 システム概要

システムの概要を図 1 に示す．

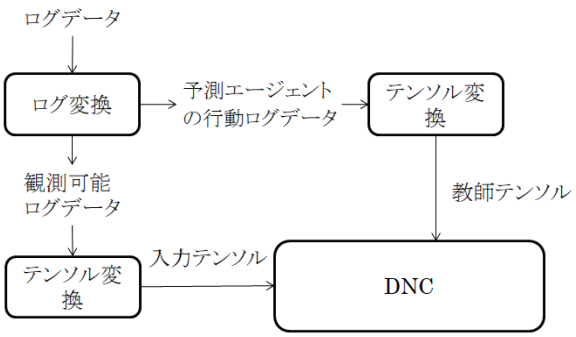

図 1: 構築した学習システムの概要

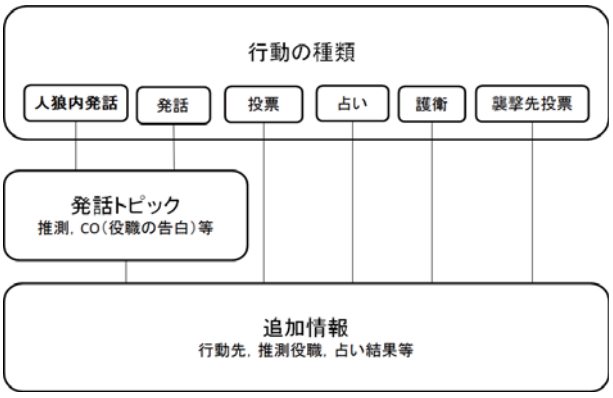

図 2: ログデータの概略

3.3 ログデータ

第 3 回人狼知能大会時でのプロトコル (人狼知能プロトコル ver2.01) を採用した [鳥海]．各エージェントについて順番に行動が回っていき，その時に行える 1 つの行動が 1 行に対応する．ログは階層構造になっており，図 2 に示すような情報を含んでいる．

3.4 ログ変換

ログデータには，すべてのプレイヤーのすべての行動，処刑や護衛先の役職などの非公開情報などプレイ中は手に入らない多くの情報が含まれている．よってこのログから適切な情報を抽出する必要がある．今回扱うタスクはあるプレイヤーの行動をそのプレイヤーの観測情報から予測できるかというものであるので，入力としては，あるプレイヤーから観測できる周りの情報を抽出する必要がある．Differentiable Neural Computer には出力を再び入力に入れるような再帰構造があるため，自分のこれまでの行動は情報として入力されるがログデータから明示的に入力として与えることはない．また，出力，つまり教師データとしてはそのプレイヤーの行動情報を抽出する必要がある．よって，まずランダムに一人のプレイヤーを選択し，そのプレイヤーの視点で獲得できる情報を選別する．プレイヤーを定めたのち，ログデータに含まれる情報は 5 種類に分類される．

1. 自分の行動情報
2. 他人の行動情報（観測可能）
3. 他人の行動情報（観測不可能）
4. システムからの情報（観測可能）
5. システムからの情報（観測不可能）

自分の行動情報とは選択したプレイヤーの投票先や護衛先，発言内容といった内容である．他人の行動（観測可能）というのは，選択したプレイヤー以外の発言内容や投票先といった情報で，プレイ中に選択したプレイヤーが知ることが可能な他人の行動である．他人の行動（観測不可能）というのは，選択したプレイヤー以外の護衛先，占い先など，本人のみに告げられ，他人（選択したプレイヤーを含む）には知らされない行動である．システムからの情報（観測可能）というのは選択したプレイヤーの役職や日付，ゲームの進行などのシステム側から与えられる情報のうち観測が可能なもののことである．人間同士で

の人狼ゲームでは全体や自分に向けてゲームマスターが公開した情報のことである．システムからの情報（観測不可能）というのは選択したプレイヤー以外の役職，自分以外の占い師の占い結果などのゲームマスター，あるいはシステム側から与えられる情報であるが，選択したプレイヤーには公開されない情報のことである．以上のように分類を行ったとき，学習モデルへの入力として他人の行動（観測可能）とシステムからの情報（観測可能）を抽出する．また，学習モデルへの教師データとして自分の行動情報を抽出する．

3.5 Differentiable Neural Computer

今回学習モデルとして Differentiable Neural Computer (DNC)[Graves 16] を利用する．DNC は通常の neural network とは異なり，外部メモリを持つ点が特徴である．外部メモリへの入出力制御まで学習することが可能であり，今までのネットワークでは扱えないような地下鉄の路線図から最短経路を探索するタスクやパズルの解法などの複雑なものを学習することが可能であると報告されている [Graves 16]．

4. 実験内容

4.1 実験仮説

1. 1 種類のエージェントを 15 体用意し，そのエージェントで 15 人人狼ゲームを行ったとき，選択した一人のエージェントの行動が予測できる．

2. 第 3 回人狼知能大会決勝のログデータについて，選択した一人のエージェントの行動が予測できる．

4.2 モデルの入出力

ログデータからシステムのログ変換機能を使って，入力データと教師データを作成する．具体的な入力の形状として，(バッチサイズ) ×（ログデータの行数）×（ 1 行当たりのベクトル長) の 3 次元テンソルを入力する．今回はミニバッチ学習を用いたため，並列に学習できるようなネットワーク構造をしている．そのため，入力はバッチサイズとして指定した数のゲームを変換したテンソルを入力とする．また，ログデータの行数の次元を系列として扱うようになっている．つまり，1 行を一つのベクトルに変換し，そのベクトルを系列として扱う．出力も同様のものを要求する．出力が不要の行については教師データはすべて 0 になっているが，誤差関数の計算には含まれない．よってどのような出力をしても誤差に影響されないものとする．

4.3 評価方法

行動が正しく予測できたかの基準について述べる．各行動ごとに行動をとるのに必要な情報が一致しているかを予測できたかの判断基準として定めた．行動の種類はゲーム進行の都合によって決まり，ゲームマスター側からの指定があってプレイヤー側では選択出来ないため，一致しているかどうかを行動の評価基準には含めない．

4.4 実験

以下に示すログＡとログＢのそれぞれの全体の予測結果の一致率について同一のハイパーパラメータを用いて学習させて比較する．ただし学習に用いたデータ数は異なる．

ログＡ： 人狼知能プラットフォーム 0.4.9 に付属するサンプルエージェントを 15 体用意し，15 人人狼を行った場合のログデータ

ログＢ： 第 3 回人狼知能大会決勝でのログデータ

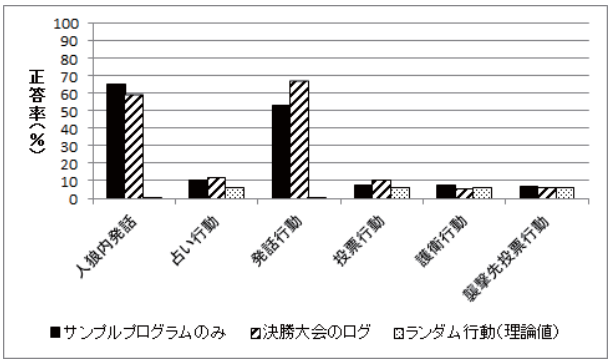

図 3: 行動の種類ごとの正答率の比較

上記のログＡとログＢでの各行動ごとの予測結果の一致率についても同一のハイパーパラメータを用いて学習させて比較する．また各行動ごとの難易度の基準として，行動を行う場合に必要な選択肢を，すべてランダムに予測した場合の一致率についての理論値との比較も行う．

4.5 データ数

サンプルプログラムのみ： 学習データ 1350, テストデータ 150

決勝大会のログ： 学習データ 517500, テストデータ 37500

5. 実験結果

学習結果として表 1 図 3 のような結果が得られた．ランダム行動は，行動に必要な情報をすべてランダムに予測した場合の正答率について理論値を計算したものである．

表 1: 各ログの予測一致率

	サンプルプログラムのみ	決勝大会のログ
一致率	49.15%	62.68%

6. 考察

6.1 実験結果に対する考察

まず，全体の正答率について考える．

タスクの難易度としてサンプルプログラムのみの行動予測と決勝大会のログの行動予測では，サンプルプログラムのみの行動予測のほうが簡単であるはずである．理由として，1 つのエージェントにつき一つのアルゴリズムが存在するため，決勝大会のログには複数の行動指針が含まれることが挙げられる．今回のシステムには，人狼ゲームでの共通の思考要素を学習することを期待していたため，エージェントの種類の情報を与えていない．そのため，複数の行動指針が含まれるような場合，あるエージェントにとってはこの行動が正解であるが，ほかのエージェントにとっては違う行動が正解であるといったような複数の正解パターンが生まれる場合もある．

今回の結果を見てみるとサンプルプログラムのみ 49.15%，決勝大会のログ 62.68%となっている．ここから，考えていたタスクの難易度とは逆の結果が得られた．この結果について以下の 3 つの原因が考えられる．

1. 学習データ数が大きく違うこと

2. バッチサイズが小さすぎる

3. 学習の打ち切りタイミングが固定である

学習データが多くなれば，その分様々な場面のデータが含まれるため，より精度よくモデルを近似できると考えられる．今回は少ないデータでも過学習を起こしていなかった．そのため，大きな問題はないと考えていたが，この影響を無視することはできない．また，時間と資源の都合（この実験は GPU は使わず学習した）上バッチサイズを小さくしていた．しかしクロスエントロピー誤差が大きく増減していたため打ち切ったタイミング次第で正答率も大きく変わると考えられる．その為，学習の打ち切りタイミングが固定であることは問題となっていると考えられる．

次に，各行動ごとの予測精度の比較を行う．どちらのログから学習した場合でも，発話行動と人狼内発話の 2 つの行動についての予測精度は，ランダム予測と比べて非常に高い．一方でほかの行動に関してはランダム行動時とほぼ変わりがない．人狼内発話と発話行動はともに会話にかかわる行動であり，どちらの行動も “日付，行の種類（人狼内発話か発話行動），発話 ID，ターン番号，発話エージェント番号，発話内容” といった同一のプロトコルで扱われる．よって，人狼内発話と発話行動に用いるプロトコル（以下，発話プロトコル）に関しては学習ができたと考えられる．逆にそれ以外は学習できていないと考えられる．これは全行動に占める割合に着目すると説明できる．発話プロトコルを用いる行動数はサンプルプログラムのみ，決勝大会のログどちらでも全体の 9 割以上である．多く出現する行動をまず学習するというのは誤差を小さくするうえで妥当であると考えられる．よって，まず発話プロトコルを十分に学習し，そののちほかの行動を学習すると考えられる．今回は学習を規定回数で止めているため，発話プロトコル以外が十分に学習されなかったのではないかと考える．よって，より規模の大きい学習が必要であるだろう．そこで現在，GPU マシンでバッチサイズを 2 から 64 に大きくして学習中である．学習途中の結果として今回とほとんど変わらない結果しか得られていないが，クロスエントロピー誤差がほぼ線形に下がり続けているため学習量が十分ではないと考えられる．一方で学習ステップ自体は今回の実験より少ないもののバッチサイズが非常に大きくなり，学習時間も数倍になってもほとんど結果が向上しないことを考えると学習の規模では大きな改善が得られないと予想される．

発話行動以外の行動の一致率が低い原因の一つについて，図 2 に示す行動の種類についての情報の内，自分の行動に関するものを，入力に含めなかったことが考えられる．行動の種類はプレイヤーが自由に選択できるものではなく，ゲームの進行の都合によって固定される．そのため，本来学習対象ではない．そこでその情報を入力として与えないのは適切ではない．これは通常の人狼ゲームでは会話の時間が終わり，投票の時間が始まったことを指示されないこと等に相当する．周りの人が投票を始めたことで，投票しなければならないことに気づくかもしれないが，指示がない分だけ判断が難しくなることは明らかである．自分の行動の種類が判断しづらくなったため，すべての行動を発話行動だとみなして行動し発話行動以外の確率が下がったのではないかと考える．

6.2 今後の展望

まず必要となるのが会話プロトコル以外の学習可能性や会話プロトコルかどうかの判断に対する学習可能性を調べるため，クロスエントロピー誤差が収束する程度までは学習させる必要がある．次に，自分の行動の種類をシステムからの情報に含めて，行動時の入力に含める必要があると考える．

また，今回発話の学習可能性が示せたため，それを利用したエージェントとして，勝利エージェントのみを学習させることで強い発話行動を持ったエージェントが作れるのではないかと考えている．もし，会話プロトコル以外も学習できることが示せれば，目標としていた行動すべてを学習によって行うエージェントが作成できるのではないかと期待している．

7. 結言

今回，DNC を用いた学習によって，発話にかかわる行動を学習できることを示した．また，今回は発話以外の行動に関しては学習可能であることを示すことができなかった．それには学習規模と入力情報の選択の 2 つの要因が考えられる．学習規模については現在調査中で，ある程度の大きくしても結果は向上しなかったが，クロスエントロピー誤差が下がり続けているため，学習を続ける必要がある．今後，入力情報の選択に関しても調査していく必要がある．また，勝利エージェントのみによる学習等を用いて，目標である行動すべてを学習によって行う人狼知能エージェントの作成に向けて取り組んでいきたい．

参考文献

[Graves 16] Graves, A., Wayne, G., Reynolds, M., Harley, T., Danihelka, I., Grabska-Barwińska, A., Colmenarejo, S. G., Grefenstette, E., Ramalho, T., Agapiou, J., et al.: Hybrid computing using a neural network with dynamic external memory, *Nature*, Vol. 538, No. 7626, p. 471 (2016)

[Silver 16] 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, pp. 484–489 (2016)

[篠田 14] 篠田孝祐, 鳥海不二夫, 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, 人工知能学会全国大会論文集, Vol. 28, pp. 1–3 (2014)

[鳥海] 鳥海 不二夫, 稲葉 通将, 大澤 博隆, 片上 大輔, 松原 仁, 狩野 芳伸, 大槻 恭士, 園田 亜斗夢：人狼知能プロジェクト, http://aiwolf.org/

[鳥海 14] 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 他：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, Vol. 2014, pp. 127–132 (2014)

[鳥海 16] 鳥海不二夫, 片上大輔, 大澤博隆, 稲葉通将, 篠田孝祐, 狩野芳伸：人狼知能　だます・見破る・説得する人狼知能, 森北出版 (2016)