JAIST Repository
https://dspace.jaist.ac.jp/

Title	人間プレイヤを活躍させる協力型ゲームの味方AI
Author(s)	板東，宏和；池田，心；Hsueh，Chu-Hsuan
Citation	情報処理学会第49回GI研究発表会，2023-3
Issue Date	2023-03-17
Type	Conference Paper
Text version	publisher
URL	http://hdl.handle.net/10119/18233
Rights	社団法人 情報処理学会，板東 宏和，池田 心，Hsueh Chu-Hsuan，情報処理学会第49回GI研究発表会，2023-3，2023．ここに掲載した著作物の利用に関する注意 本著作物の著作権は情報処理学会に帰属します。本著作物は著作権者である情報処理学会の許可のもとに掲載するものです。ご利用に当たっては「著作権法」ならびに「情報処理学会倫理綱領」に従うことをお願いいたします。 Notice for the use of this material: The copyright of this material is retained by the Information Processing Society of Japan (IPSJ). This material is published on this web site with the agreement of the author (s) and the IPSJ. Please be complied with Copyright Law of Japan and the Code of Ethics of the IPSJ if any users wish to reproduce, make derivative work, distribute or make available to the public any part or whole thereof. All Rights Reserved, Copyright (C) Information Processing Society of Japan.
Description	情報処理学会第49回GI研究発表会，2023-3，明治大学

JAIST
JAPAN ADVANCED INSTITUTE OF SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Vol.2023-GI-49 No.3
2023/3/17

人間プレイヤを活躍させる協力型ゲームの味方 AI

板東 宏和[1,a)]　池田 心[1,b)]　Hsueh Chu-Hsuan[1,c)]

概要：マルチプレイヤゲームでは，プレイヤ同士の協力が求められる．そのため，協力することによって強いチームとなる味方 AI の研究はこれまで行われてきた． しかし，このようなゲームの目的を達成することを重視した行動は，時に味方の意図を無視している場合がある．人間プレイヤはたとえ目的を達成しても，自分の意図が理解されず活躍できないと感じれば不満を抱く可能性がある．翻って，意図を理解して活躍させる行動は，人間プレイヤを楽しませる味方 AI として重要である． 本研究では，人間プレイヤを活躍させるための協力に焦点を当て，意図を持ったプレイヤを表現したエージェントと，そのエージェントをサポートする味方 AI を作成し，意図推定器によって味方 AI を切り替える手法を提案した．評価実験により，特定の意図を持ったプレイヤエージェントをサポートできることを明らかにした．また，意図推定器が正しく意図を推定することで，プレイヤの持つ意図に沿ったサポートが可能になることを示した．

キーワード：協力型ゲーム，活躍させる，意図推定，マルチエージェント強化学習

Teammate AI in cooperative games letting human players play active roles

1. はじめに

昨今，人工知能（AI）技術は大きな発展を遂げており様々な用途で利用されている．車の自動運転や動画像の生成など，AI 技術は人間の代わりに用いられることも多い．中でもディジタルゲームでは，人間の代替として良いゲーム AI には，人間と並ぶ強さが求められる．そのため，強いゲーム AI は研究の対象となっており，Agent57[1] はディジタルゲーム atari2600 の 57 ゲームすべてで，人間プレイヤを上回る得点を記録している．このように，ゲーム AI の研究は強さを求める目的において十分な成果を上げている．

ディジタルゲームの中でも協力型マルチプレイヤゲームでは，人間同士でチームを組めないような場合に味方プレイヤとしてゲーム AI が必要である．そのため，このようなゲームジャンルではゲーム AI と人間プレイヤとの協力が重要となる．協力を行う上で求められる行動は様々あり，主目的達成のための協調行動，プレイヤの価値観に合わせた協調行動，人間プレイヤらしい行動などが挙げられる．

一方で，人間にとってこれらの協調行動が最も好ましいとは限らない．仮にチームとして勝てたとしても，自分が何らかの意味で “活躍” できなければ面白く感じられないだろう．そこで本論文は，多少主目的の達成を犠牲にしてでも人間プレイヤを引き立て活躍させる味方 AI の作成を目指す．これは，人間の意図を尊重してそれを支援する協調行動を行うことで実現すると考える．

2. 関連研究

先行研究において，協調行動は主に 2 種類の目的に区別される：主目的達成のためと副目的達成のため．

主目的達成のための協調行動は，ゲームにおける利得を最大化するための行動である．副目的達成のための協調行動は，ゲームの利得とは異なる価値観を重視した行動である．たとえ主目的達成のための協調行動であっても，発生が困難なものは多い．これは，味方が上級者であるか初級者であるか，ゲーム AI であるかという前提によって最適な協調行動が異なるためである．

主目的達成のために協力するゲーム AI の研究として，DeepMind 社が発表した Fictitious Co-Play(FCP) では，

1 北陸先端科学技術大学院大学
Japan Advanced Institute of Science and Technology
a) s2110139@jaist.ac.jp
b) kokolo@jaist.ac.jp
c) hsuehch@jaist.ac.jp

ⓒ2023 Information Processing Society of Japan

様々な人間プレイヤに好かれる協調行動が実現されている [2]．この研究では，作成されたエージェントは人間プレイヤのプレイデータを用いず，エージェント自身の学習途中のモデルを味方とする．このようなモデルと別のパラメータを持ったエージェントとの組み合わせにより，多くのレベルの人間プレイヤに適応した．

また，副目的達成のために協力するゲーム AI の研究として，和田らによるプレイヤの価値観を学習したものがある [3]．この研究では，人間プレイヤの重視している目的を副目的として効用関数で表現することに着目している．プレイヤごとにとりうる各行動がもたらす結果をランダムシミュレーションの平均帰結として求めることで，効用関数を学習し，その価値観に合わせた行動をとるゲーム AI を作成した．これにより，コマンド選択式の RPG ゲームにおいて，特定の効用重みをもった関数よりも自然な挙動が得られることを示した．

3. 対象とするゲーム

3.1 Unity ML-Agents

味方と協力するゲーム AI の研究は，多くのマルチプレイヤゲームを用いて行われてきた．これに伴って，研究用のプラットフォームの開発も多く存在する．昨今では，協力型ゲームのテストベッドとして様々なライブラリが公開されている [4], [5]．Unity Technologies 社が提供している Unity ML-Agents もこのような 1 つで，これは Unity を用いて機械学習を行う環境のオープンライブラリである [6]．ML-Agents は，Python API を用いた学習が可能であり，強化学習や模倣学習をはじめとした様々なアルゴリズムが実装されていること，また多数の協力型ゲームがテストベッドとして提供されていることから本研究で利用する．

本研究では，ML-Agents で提供されているサンプル環境である DungeonEscape を用いる．このゲームは，敵キャラクタを環境の一部とみなせば完全協力型マルチプレイヤゲームである．ただし，DungeonEscape は非常に単純なゲームルールであるため，人間プレイヤの意図に基づいた行動が発生しにくい恐れがある．そこで，本研究ではこのゲームに変更を加えた拡張ルールを提案する．

3.2 拡張版 DungeonEscape

DungeonEscape からの変更部分を**表 1** に示す．DungeonEscape では，各エージェントの向いている方向のみを観測する部分観測であるため，エージェントは盤面全体の真の状態を得ることができない．この環境で味方エージェントの意図を汲み取る場合，味方エージェントを観測し続ける必要がある．ゲームをクリアすることを考えると，ドラゴンも観測していなければならないため，絶えず周囲を観測する首振り行動を取らなければ協調行動が行いにくいことになる．これは煩雑であり，協調行動を主眼においた本研究にとっては都合が悪いため，観測範囲をエージェントの前方から全方位への変更した．

表 1 DungeonEscape からの変更点
Table 1 Some modified points from DungeonEscape

要素	DungeonEscape	拡張版 DungeonEscape
観測方法	部分的	全体[*1]
鍵の位置	ドラゴンの位置	マップの中央
ドラゴン死亡時の挙動	接触エージェントの死亡	周辺エージェントの死亡
制限時間	25000 フレーム	6000 フレーム

また，鍵のドロップ位置とドラゴンの死亡条件を変更した．本来ドラゴンを倒した位置から入手できる鍵を，ドラゴンを倒した後マップ中央から鍵が入手できるようにすることで，「素早いクリアのために鍵の発生する場所で待ち伏せする」などエージェントの行動が多様化すると考えたためである．さらに，ドラゴンの死亡時の挙動として爆発を追加した．爆発とは，ドラゴンを中心とした一定範囲内にいるエージェントに攻撃するものである．エージェントはこの攻撃に触れると死亡するため，一斉に向かって行く行動では死亡するリスクがあり味方の行動を見た上で自身の行動を決定するといった協調行動が発生すると考えた．

最後の変更点は，制限時間である．DungeonEscape では，学習を円滑に行うために制限時間は存在する．しかし，この学習時間は平均ゲームクリア時間の 100 倍以上に相当し，人間がゲームをプレイする上で実質ないものとして扱うことができる．従来の設定のままでは，ドラゴンを倒した後脱出までのモチベーションが存在しないことから，ゲーム全体に制限時間を設定した．

以上の拡張版 DungeonEscape を本研究環境として扱う．また，詳しいゲームプレイの手順を**図 1** に示す．

4. 提案手法

本研究の目的は，協力型マルチプレイヤゲームで人間プレイヤを活躍させるための協力をおこなうことである．提案するアプローチの全体像を**図 2** に示す．

まず，5 章では拡張版 DungeonEscape で想定される意図を列挙し，意図を活躍させる行動について考察する．6 章でその意図に基づく行動をとる意図エージェントを学習する．本手法ではコストや時間の面から，被験者データを使用せず人間の代わりとなる意図エージェントを用いた．7 章では，意図エージェントをペアにしてゲームをプレイすることで，意図を汲み取って活躍させる個別意図サポート AI を学習する．また，8 章で意図エージェントのゲームデータから意図推定器を作成し，推定結果から意図に合う個別意図サポート AI に行動決定を任せる統合意図サポー

[*1] エージェントを中心とした 360 度

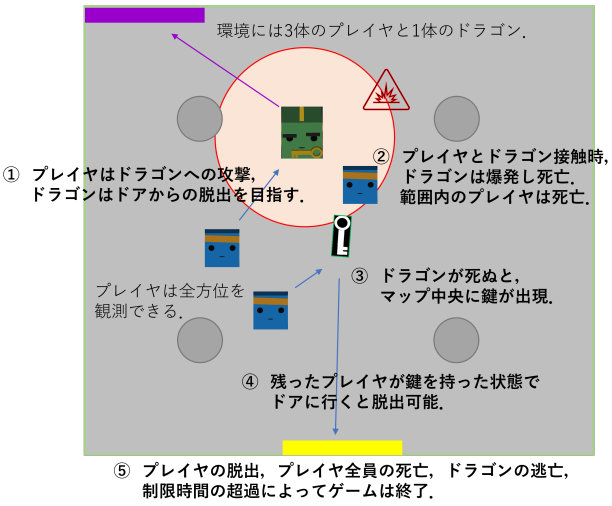

図 1 拡張版 DungeonEscape プレイの流れ
Fig. 1 How to play the expanded edition of DungeonEscape

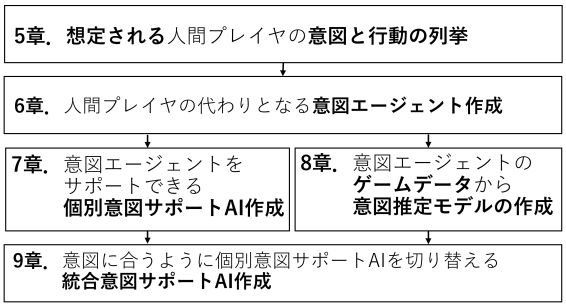

図 2 アプローチの全体像
Fig. 2 Overview of the proposed method

ト AI を 9 章で実装する.

5. 対象とする意図

5.1 想定される意図の列挙

本来，人間プレイヤが持ちうる意図は無数に存在しそれらすべてに対応することはできない．そこで，対象ゲームを拡張版 DungeonEscape に限定し，そこで発生しうる 4 つの代表的な意図を以下に挙げ考察と実装の対象とする．

(A) 攻撃志向は，“自分がドラゴンを倒したい” という意図を持っている．この意図を持つとき，ドラゴンを倒したことを活躍したと定義する．このような意図を持つ人間プレイヤの行動は，味方の初期位置に関わらず，自身がドラゴンに向かって行き接触するなどである．

(B) 生存志向は，“自分は死亡したくない” という意図を持っている．この意図を持つとき，ゲームクリア時にそのプレイヤが生存していることを活躍したと定義する．このような意図を持つ人間プレイヤの行動は，味方の初期位置に関わらず，自身がドラゴンから離れる，もしくはドラゴンに向かって行くが接触はしないものである．

(C) 効率志向は，“出来るだけ効率よくゲームをクリアしたい” という意図を持っている．この意図を持つとき，ゲームのクリアタイムが早くなることを活躍したと定義する．このような意図を持つ人間プレイヤの行動は，味方と初期配置ごとの役割分担を行うことである．

(D) 仲間志向は “味方プレイヤの行動に合わせるように協力したい” という意図を持っている．この意図は，自分が味方に協力している状況を楽しむものであり，他のプレイヤの行動を見て合わせるように協調ができた場合に活躍を感じられる．(C) と (D) の表出する行動は似ているが，(D) の場合は仲間のとった行動が主目的達成にとって不適切であったとしてもそれをサポートするという点で異なる．

以上のことから，A から D のような意図に基づく行動は，このようなゲームで頻繁に発生しうるものであると考える．しかし，D は人間プレイヤから見て「仲間を活躍させたい」ということであり，本研究でやらせたい挙動そのものである．これは，より高度な意図ないし挙動であって「仲間を活躍させたいプレイヤを活躍させる」ために AI エージェントが何をすべきかは難しい問題である．そこで，本研究では (D) は取り扱わず (A) から (C) のみを対象とする．

表 2 それぞれの意図をサポートする行動
Table 2 Actions to support each intention

意図	活躍させたいプレイヤの行動	サポート行動
攻撃志向	ドラゴンへの接近	ドラゴンへの攻撃を譲る
生存志向	ドラゴンから離れる	ドラゴンへ攻撃する
効率志向	味方の位置よって変化	味方の位置によって変化

5.2 意図に沿ったサポート

前節では対象とする意図について列挙した．この節では，それぞれの意図を活躍させる行動について述べる．表 2 は，各意図とサポート行動を対応させたものである．

攻撃志向プレイヤを活躍させるためのエージェントを攻撃サポート AI と呼び，ドラゴンに近づかず攻撃を譲る行動をとる必要がある．

生存志向プレイヤを活躍させるためのエージェントを生存サポート AI と呼び，ドラゴンに近づく行動をとる必要がある．自分が犠牲となってドラゴンを倒すことで，意図と異なる結果になることを防ぐ．

効率志向プレイヤを活躍させるためのエージェントを効率サポート AI と呼び，味方の初期位置によって適切な行動をとる必要がある．具体的には，効率志向エージェントがドラゴンへ近づく行動をとった時はマップ中央で鍵を入手するように行動し，鍵の出現場所で待機する行動をとったときはドラゴンへ攻撃するように行動すればよい．このように，味方の初期位置によってどちらの行動をとるか判断しているのであれば，効率サポート AI は同じ判断基準を持つことでこれに対応できる．したがって，効率サポート AI と効率志向エージェントの挙動は同じものであると想定できる．

ⓒ2023 Information Processing Society of Japan

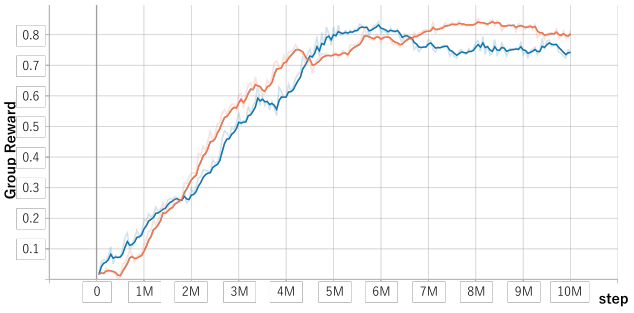

図 3 デフォルトエージェントの学習曲線
Fig. 3 Learning curve for default agent

6. 意図エージェントの作成と評価実験

6.1 意図を持たないエージェント

本節では，特別な意図や特徴を持たない「デフォルトエージェント」を学習する実験を行う．ここで学習したモデルは，次節で述べる意図エージェントとの比較のために用いる．実験目的は，マルチエージェント強化学習を用いて協力してゲームをクリアするという挙動が発生することの確認である．なお，ゲームルールは図 1 の通りである．

プレイヤは，自身の位置から全方位を不透過センサによって観測し，前後左右への移動，左右の方向転換ができる．不透明センサは｛壁，プレイヤ，ドラゴン，鍵，ドア，ポータル｝との衝突判定と距離を観測する．この環境でマルチエージェント強化学習手法 MA-POCA[7] を用いて 3 体のプレイヤを学習する．報酬はエージェント全体に対する報酬 R_g と各エージェントに対する個別報酬 R_i として与えられる．R_g は，より早いゲームクリアを望む全体報酬である．また，R_i は，行動変更数にペナルティを与えることで，人間らしくない不自然で細かい機械的な挙動を抑制する目的がある．R_g と R_i は，ゲームクリアまでにかかった時間 f とプレイヤの行動変更回数 c によって次のように表される．

$$R_g = (0.999)^f \tag{1}$$
$$R_i = -0.001c$$

以上の設定で 10M ステップ学習を行ったときの平均報酬を**図 3** に示す．学習には 5 時間程度を要した．学習は 2 セット行い，図中の曲線（赤，青）は移動平均をとったものを表している．5M ステップで学習時全体報酬はおよそ 0.8 になることが確認できた．どちらも平均ゲームクリア率は 0.9，平均エピソード長は 140 フレーム（2.3 秒）となった．

6.2 意図エージェントの学習方法

ここでは，3 種類の意図エージェントを学習する．意図エージェントの行動は，意図を持った人間の振る舞いと似ていることが望ましい．各意図は，与える報酬の変更によってのみ定義する．このようにして，攻撃志向（式 2），生存志向（式 3），効率志向（式 4）の行動に合わせた報酬をそれぞれ以下のように定めた．

$$R_g(\text{attack}) = (0.999)^f \tag{2}$$
$$R_i(\text{attack}) = \begin{cases} -0.001c + 0.3 \text{ (killed dragon)} \\ -0.001c - 0.1 \text{ (didn't kill dragon)} \end{cases}$$

$$R_g(\text{escape}) = (0.999)^f \tag{3}$$
$$R_i(\text{escape}) = \begin{cases} -0.001c - 0.5 \text{ (dead)} \\ -0.001c \text{ (alive)} \end{cases}$$

$$R_g(\text{efficient}) = (0.99)^f \tag{4}$$

環境に存在する 3 体のエージェントはすべて同じ報酬を与えて学習を行う．攻撃，生存志向エージェントの個別報酬は，そのエージェントがドラゴンを倒したかどうかで異なる．攻撃志向エージェントの場合はドラゴンを倒せば，生存志向エージェントの場合は自分が生き残れば高い個別報酬となる．また，早めのクリアを目指すことやボタン変更数が多くないほうがよいことは共通しているため，デフォルトエージェントと同じ報酬である．効率志向エージェントの報酬は，デフォルトエージェントの報酬 R_g を基にクリア時間による変化量を強く反映させたものである．また，効率的な人間プレイヤは，焦りから操作のブレや多少の無駄な行動を行うと考える．そこで，ボタン変更による負の報酬を無くすことによって焦りを表現する．与える報酬以外の学習時の設定は 6.1 節の実験のものと同じにして学習を行った．

6.3 意図エージェントの性能評価

各意図エージェントは 5M ステップ学習を行った．また，デフォルトエージェント 2 体をペアにしたゲームを 2000 試行行うことで評価した．例えば攻撃志向モデルでは，攻撃志向エージェント 1 体，デフォルトエージェント 2 体のチームでゲームを行ったときの攻撃志向の結果を示している．なお，比較対象としてデフォルトエージェントがあるが，こちらもデフォルトエージェント 2 体をペアにしたもの（3 体とも同じ）である．各意図エージェントの学習結果を**表 3** に示す．また，評価項目は次の通りである．

- クリア率：終了したゲーム中のクリア割合
- 攻撃率：クリアしたゲーム中の評価対象のエージェントが攻撃した割合
- ボタン変更数：終了したゲーム中の 1 ゲーム中に対象エージェントが操作を変更した回数
- エピソード長：クリアしたゲームと時間切れとなったものについて，1 ゲームにかかった時間

表 3 意図エージェントのシミュレーション結果
Table 3 Simulation results for intention agent

評価対象 AI	クリア率	攻撃率	ボタン変更数	エピソード長
デフォルト	0.916	0.331	6.1	141.1
攻撃志向	0.941	0.408	6.8	139.3
生存志向	0.871	0.012	47.9	230.5
効率志向	0.906	0.317	11.5	124.7

攻撃志向エージェントでは，クリア率は，0.941 とデフォルトエージェントと比較して 0.025 の増加があった．また，攻撃率は 0.408 であり，4 割以上攻撃を行っていることが確認できる．これは，攻撃志向エージェントが積極的にドラゴンへ攻撃していることを示す．

生存志向エージェントでは，攻撃率は 0.012 であり，ほとんど攻撃しなかったことが確認できる．ボタン変更数は大きく増加しており，ドラゴンから離れるために細かな動きが多かったことは興味深い結果となった．さらに，エピソード長は非常に大きな値をとった．これは，ドラゴンに接近しないことを優先する行動が，全体のクリアタイムを大きく低下させるという結果を示す．

効率志向エージェントでは，ボタン変更数は 11.5 となり，デフォルトエージェントの 2 倍ボタン操作を変更していることから，デフォルトエージェントと似ている行動をとりながら細かく動いていることが示された．加えて，エピソード長は 124.7 であり，デフォルトエージェントよりも早くゲームをクリアしていることが分かった．これは，効率志向の人間プレイヤの焦りと効率的なゲームプレイを表現している．

以上の結果から，攻撃志向エージェントでは攻撃率が高く，生存志向エージェントでは攻撃率が低い，また効率志向エージェントではボタン変更数の増加とクリアタイムの短縮が見られた．これは，概ね狙い通りであり，意図を表現したエージェントが作成できたと考える．ただし，これは意図から定義した行動を意図エージェントが獲得したものであり，人間の意図を正確に表現しているとは限らない．

7. 意図サポート AI の作成と評価実験

7.1 意図サポート AI の学習方法

意図サポート AI は，意図エージェントをペアにして学習を行う．例えば，攻撃サポート AI の学習時に攻撃志向エージェントは味方として行動するが，方策の改善が行われるのは 2 体のサポート AI のみである．

攻撃サポート AI は，攻撃志向がドラゴンを倒した時に正の報酬を受け取り，倒せなければ負の報酬を受け取る．これは，自分がドラゴンをできるだけ倒さないことを狙った報酬で，式 2 とは反対の志向性を持っている．反対に，生存サポート AI は生存志向が生き残ったときに正の報酬を受け取る．これは，生存志向をできるだけ死なせないようにする報酬で，式 3 を共有したものと言える．効率サポート AI は，結果的に効率志向の学習と同じことであるため，新たな学習は行わずに効率サポート AI を効率志向エージェントとして扱うことにする．与える報酬以外の学習時の設定は，6.1 節の実験のものと同じにして学習を行った．

7.2 意図サポート AI の性能評価

各意図サポート AI の作成では，5M ステップ学習を行った．また，意図エージェント 1 体をペアにしたゲームをそれぞれ 2000 試合行うことで性能を評価した．例えば，攻撃サポートの評価では，攻撃志向エージェント 1 体，攻撃サポート AI 2 体のチームでゲームを行ったときの攻撃志向の結果に注目する．評価項目は前節と同じものである．結果とデフォルトエージェントがペアのときの比較を**表 4** にまとめる．

攻撃サポート AI の評価では，攻撃志向 1 体にデフォルトエージェント 2 体をペアとしたときのクリア率 0.941 から，攻撃サポート AI 2 体をペアにすると 0.821 となり低下が見られた．また，攻撃率は，0.408 から 0.896 と大きく増加した．エピソード長は 139.3 から 173.4 フレームと増加しているが，これは攻撃サポート AI がドラゴンから離れる行動をとっていたためであると考える．

また，生存サポート AI の評価では，生存志向 1 体にデフォルトエージェント 2 体をペアにしたときのクリア率 0.871 から，生存サポート AI 2 体をペアにすると 0.860 となり大きな変化はなかった．攻撃率は 0.012 から 0.005 となり，ドラゴンへ攻撃することがほとんどなかったことが分かる．これは，生存サポート AI が生存志向エージェントに合わせた積極的なドラゴンへの攻撃行動を学習したためであると考える．ボタン変更数は 47.9 から 55.5 とわずかに増加しているが，これは生存志向エージェントの活動時間に起因するものである．

最後に，効率サポート AI の評価では，効率志向 1 体にデフォルトエージェント 2 体をペアにしたときのエピソード長 124.7 フレームから，効率サポート AI 3 体のときには 113.0 フレームと短縮されていた．これは，効率的にクリアしたいという意図をサポートすることができたと考える．

以上の結果から，味方がデフォルトエージェントであるときと比較したとき，攻撃サポート AI のときは攻撃率が上昇し，生存サポート AI のときは攻撃率が低下した．効率サポート AI のときもクリアタイムが短縮され，3 種類の意図をサポートする行動が見られた．

8. 意図推定器の提案

8.1 推定器の学習方法

ドラゴンと対象プレイヤとの距離，プレイヤとドラゴンの位置情報を入力データとする．意図によってそれぞれとる行動が異なり，特にドラゴンに近付くかどうかが重要であるためこのような入力を用いた．また，意図推定器は

表 4 意図サポート AI シミュレーション結果の比較
Table 4 Comparing simulation results for support AI

サポート AI	クリア率	攻撃志向エージェントの攻撃率	攻撃志向エージェントのボタン変更数	エピソード長
デフォルト	0.941	0.408	6.8	139.3
攻撃サポート AI	0.821	0.896	6.2	173.4
サポート AI	クリア率	生存志向エージェントの攻撃率	生存志向エージェントのボタン変更数	エピソード長
デフォルト	0.871	0.012	47.9	230.5
生存サポート AI	0.860	0.005	55.5	251.8
サポート AI	クリア率	効率志向エージェントの攻撃率	効率志向エージェントのボタン変更数	エピソード長
デフォルト	0.906	0.317	11.5	124.7
効率サポート AI	0.910	0.323	10.6	113.0

ゲーム中に対象プレイヤの意図を推定する必要があり，かつ推定した段階でゲームが終了していないことが求められる．そこで，デフォルトエージェントの平均エピソード長が 141 フレームであることを踏まえてゲーム開始から 30 フレーム分の情報を取得する．

出力データは意図である．対象とした情報では効率志向の意図を判断することが難しいと考え，今回は攻撃志向と生存志向の 2 種類の意図を出力として扱う．各意図を one-hot エンコーディングしたものを教師データとしてラベリングを行った．これは，攻撃志向を (1,0)，生存志向を (0,1) として扱うものである．今回は効率志向の判断を行っていないが，入力特徴量にボタン変更数を加えるなどの工夫で効率志向を含めた意図の判断が可能になると考えている．

ネットワーク構造は全結合モデルを用いた．150 次元のデータを入力として全結合層に入れ，これを 2 次元データとして出力する．なお，全結合層の活性化関数には ReLU を用いた．本研究では単純なネットワークを用いたが，入力情報量が増えて複雑になる場合は時系列データの扱いを得意とするモデルの利用も必要になると考える．このようにして，攻撃・生存志向それぞれ 2000 ゲームを取得して，これを学習 2800，テスト 1200 と分割して使用する．

8.2 意図推定器の性能評価

前述のとおり学習用に用いたデータ数は攻撃，生存志向で 1400×2，推論用に用いたデータ数は攻撃，生存志向で 600×2 である．攻撃志向は 600 回中 531 回推定が成功，生存志向は 555 回成功となり，正解率は 0.9 程度となった．攻撃，生存志向ともに概ね学習できていると判断し，これを統合意図サポート AI として組み込む．

9. 統合意図サポート AI の実装

9.1 ゲーム内での意図推定手法

ゲーム中に行動から意図を推定するためには，いくつかの条件が求められる．まず，意図に基づいた行動が発生するタイミングが問題となる．人間プレイヤはゲーム中に意図が変わることがあるが，このタイミングは一定ではなく，規則性も明らかになっていない．しかし本環境では，ゲームの平均エピソード長が 140 フレーム程度である点から，1 ゲーム中にプレイヤの意図が変化することはあまりないと考える．また，人間はゲーム開始時から何かしらの意図を持っておりその意図に沿って行動すると考える．故に，本ゲームでは開始時点で持っている 1 つの意図に従って行動し変化することはないと仮定して，ゲーム中一度だけ推定することで意図に沿った支援が可能になると想定する．

また，どの程度の時系列データを入力として与えれば，高い精度で推定できるのかという問題もある．前提として，ゲームが終了するより前に意図推定ができている必要がある．意図推定器の作成では，ゲーム開始から 30 フレーム分の行動を入力としていたが，人間プレイヤの場合は意図エージェントほど素早く迷いなく意図に応じた行動を取るわけではないため，30 フレームで十分な推定精度が得られるかは今後の課題である．

表 5 統合サポート AI のシミュレーション結果比較（攻撃志向，生存志向）
Table 5 Comparing simulation results for Integrated Support AI(attack, escape)

味方 AI	クリア率	攻撃率	意図推定精度
デフォルト	0.941, 0.871	0.408, 0.012	
個別意図サポート AI	0.821, 0.860	0.896, 0.005	
統合意図サポート AI	0.810, 0.856	0.774, 0.010	0.894, 0.920

9.2 統合意図サポート AI の実装結果

統合意図サポート AI によって意図に対応したサポートが可能となったかを明らかにする．性能評価のために，個別意図サポート AI をペアにしたものと統合意図サポート AI をペアにしたものをそれぞれ 2000 試合シミュレーションした．**表 5** は，活躍させたいプレイヤとして攻撃・生存志向エージェントを用いたときのシミュレーション結果である．評価項目は前節のものとほぼ同じである．意図推定精度は終了したゲームのうち，意図推定器が正しく意図を推定できた割合を指す．

表 5 から統合意図サポート AI をペアにしたとき，攻撃志向のクリア率は 0.81 で攻撃サポート AI のクリア率から大きな変化は見られなかった．攻撃率は 0.774 となり，0.896 から 14 %ほど低下した．意図推定精度は 0.89 程度であることから，攻撃率の低下は概ね妥当なものであると考える．生存志向では，統合意図サポート AI をペアにした時のクリア率は 0.856 で生存サポート AI のクリア率とほとんど変らなかった．意図推定精度は 0.92 であった．攻撃率がデフォルトの 0.012 付近になったのは，統合意図サポート AI の非活性化区間が原因であると考える．

以上により，攻撃，生存の 2 種類の意図に限った意図推定では概ね良好な結果が得られた．しかし，これはあくまで意図エージェントを人間プレイヤの代わりに用いた実験であり，より多様な挙動をとる人間プレイヤに十分対応できるかを明らかにするためには別の実験を要する．次節では，ロバスト性の評価を試みる．

9.3 ロバスト性の評価

9.3.1 評価実験

本節では，統合意図サポート AI のロバスト性を評価する．評価手法として，プレイヤが人間であることを考慮して意図エージェントにブレや判断の遅れといったノイズ行動を発生させ，統合意図サポート AI の性能がどの程度低下するのかを見ることで，人間プレイヤをペアにしたときの汎用性を確認する．

項目 (A) は，ランダムに与えられる時間 $t(0 < t < 30)$ まで動かない意図エージェントである．実際の人間は，ゲーム画面を認識するのにも，どのように行動するかを決定するのにも時間を要する．そのため，これは人間プレイヤの盤面認識の遅れを想定したものである．

項目 (B) は，開始から意図を推定するまで，確率 ϵ でランダム行動を行う意図エージェントである．人間プレイヤは，自身の位置を正しく把握して正確なボタン操作を行うことができないことがある．これは人間プレイヤの行動のブレからくるものであり，(B) はそれを表現している．今回は $\epsilon = 0.1$ として検証する．

9.3.2 実験結果

ノイズを付与した攻撃・生存志向をそれぞれ統合意図サポート AI のペアにして，各 2000 ゲームシミュレーションした．また，ノイズのない攻撃・生存志向をペアにしたものと比較を行った．**表 6** は，攻撃・生存志向エージェントにノイズを含めた場合の性能評価である．

表 6 の (A) を見ると，平均クリア率は 0.68 であり，ノイズのない攻撃・生存志向エージェントのクリア率 0.83 と比較して 0.15 ほど低下した．攻撃志向エージェントの推定精度は 9 割を超えていた一方で，生存志向エージェントの推定精度は 5 割にとどまった．故に，ドラゴンへの攻撃役がいなくなり，クリア率の低下や生存志向の攻撃率の増加につながったのだと考える．また，停止するという行動は攻撃志向だと判断されやすいことが分かった．

表 6 統合意図サポート AI のロバスト性の評価
Table 6 Evaluate the robustness of Integrated Support AI

活躍させたい対象のモデル	平均クリア率	攻撃率（攻撃，生存）	意図推定精度（攻撃，生存）
（ノイズなし）	0.833	0.774, 0.010	0.894, 0.920
(A)	0.682	0.763, 0.027	0.920, 0.519
(B)	0.855	0.739, 0.008	0.811, 0.956

表 6 の (B) を見ると，平均クリア率に大きな変化はなく，攻撃志向エージェントの攻撃率は低下がみられた．これは単純にランダム行動によるものと考えている．また，意図推定精度にも大きな変化はなかった．今回は $\epsilon = 0.1$ であったため，ランダム行動をとる確率が低く有意な差が生まれなかったと考える．

以上から，意図推定精度は下がったがクリア率に大きな低下はなく今回のような単純なノイズではさほど性能は落ちないことが明らかとなった．

9.4 人間的な評価

9.4.1 被験者実験

本節では，統合意図サポート AI の性能を人間的側面から評価する．人間プレイヤとの被験者実験を行い，人間をサポートできているか満足度は向上したかを見ることで本手法の実用性を明らかにする．環境は，本論文で論じた拡張版 DungeonEscape である．人間プレイヤは［前進，右移動，左移動，後進］の操作が可能である．ここで，意図エージェントは全方位レイセンサ観測であったが，人間プレイヤは 3 人称視点での全体観測ができるように変更している．また，人間プレイヤには，ゲーム開始前に初期状態を観測するための任意の時間を与えている．味方 AI と人間プレイヤとのチームでそれぞれ 15 回のゲームを行う．味方 AI は「デフォルトエージェント 2 体」「攻撃サポート AI 2 体」「生存サポート AI 2 体」「統合意図サポート AI 2 体」の 4 組を用意した．評価はアンケートによって行い，質問項目は「ゲーム中の考えはどちらに近いか（攻撃志向，生存志向）」「味方 AI をどれほど好ましく感じたか（5 段階評価）」である．

9.4.2 実験結果

5 人の被験者に対して実験を行った結果を述べる．**図 4** と**図 5** は，攻撃志向と生存志向の被験者がどの味方 AI を好ましく思ったかを示している．

図 4 を見ると，デフォルト AI は好ましさ平均 3.4 であることがわかる．比較して統合意図サポート AI は好ましさ平均が 4.6 であることから攻撃志向の人間にとってデフォルト AI よりも攻撃志向が好まれていることが明らかとなった．また，生存サポートは好ましさ平均が 2.8 であり攻撃志向の被験者には好ましく思われていない．

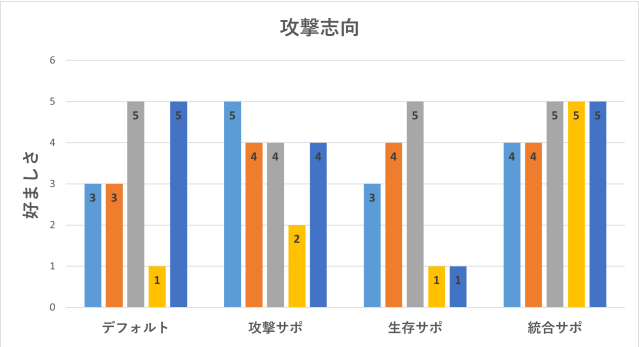

図 4 味方 AI の好ましさ比較（攻撃志向）
Fig. 4 Comparison of which teammate AI is preferred by humans. (attack-oriented)

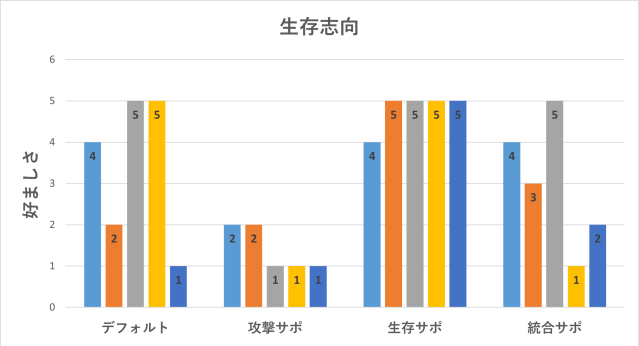

図 5 味方 AI の好ましさ比較（生存志向）
Fig. 5 Comparison of which teammate AI is preferred by humans. (escape-oriented)

図 5 を見ると，デフォルト AI は好ましさ平均 3.4 であることがわかる．比較して統合意図サポート AI は好ましさ平均が 3.0 であることから生存志向の人間にとっては統合意図サポート AI はあまり好ましく思われていないことが明らかとなった．また，攻撃サポートは好ましさ平均が 1.4 であり生存志向の被験者には好ましく思われていない．

このことから統合意図サポート AI は生存志向の人間に対しては有効に働いていないことが予測される．そこで，意図推定器による意図エージェントと人間の性能比較（**表 7**）を見ると，意図エージェントから生存志向の推定精度は大きく低下していることが分かる．これは，生存志向エージェントの挙動が生存志向の被験者を模倣できていなかったため，意図推定器が生存志向の被験者を誤って攻撃志向であると判断したことが原因である．

被験者実験の結果から，デフォルトエージェントと比較すると統合意図サポート AI は人間に好まれる傾向があるということが示された．一方で，生存志向エージェントと人間の挙動に差が見られるという課題を残した．

10. おわりに

本研究では，マルチプレイヤゲームにおいて人間プレイヤを引き立てて活躍させる味方 AI の作成を目指した．そのために，意図と協調行動が分かりやすい環境として既存のゲームを拡張した新しいゲームを提案した．このゲームで人間の意図として代表的なものを考察し，人間プレイヤの代替として作成した意図エージェントは意図に基づいた行動をとることが性能評価実験により明らかとなった．さらに，この意図エージェントをペアにしてサポート行動をとる意図サポート AI を作成した．意図サポート AI はそれぞれの意図対応できることが分かった．また，意図推定器の評価実験によって，単純なネットワークモデルでも十分に意図を分類できることが分かった．最後に，統合意図サポート AI を実装した．ロバスト性を検証することで，ある程度明確な意図があれば意図に沿ったサポートが可能であることが分かった．さらに，被験者によってデフォルトエージェントと比較すると提案手法の味方 AI は人間に好まれる傾向にあることが明らかとなった．

今後の展望として，本研究で扱った 2 種類の意図以外も扱うことで，より実用的な環境への適応を行いたい．また，意図に沿ったサポートは，人間プレイヤの満足度を向上させて楽しませるツールとして味方 AI と人間の共生を実現させるものとして期待される．

表 7 被験者の意図推定結果
Table 7 Results of human intention estimation

意図推定対象	平均意図推定精度	攻撃志向推定精度	生存志向推定精度
意図エージェント	0.905	0.887	0.923
被験者	0.717	0.929	0.490

謝辞 本研究は JSPS 科研費 JP20K12121 の助成を受けたものです．

参考文献

[1] Badia, A. P., Piot, B., Kapturowski, S., Sprechmann, P., Vitvitskyi, A., Guo, Z. D. and Blundell, C.: Agent57: Outperforming the atari human benchmark, *ICML 2020*, PMLR, pp. 507–517 (2020).
[2] Strouse, D., McKee, K., Botvinick, M., Hughes, E. and Everett, R.: Collaborating with humans without human data, *NeurIPS 2021*, Vol. 34, pp. 14502–14515 (2021).
[3] 和田堯之，佐藤直之， 池田心：少数の記録からプレイヤの価値観を機械学習するチームプレイ AI の構成，研究報告ゲーム情報学（GI）， Vol. 2015-GI-33, No. 5, pp. 1–8 (2015).
[4] Vinyals, O. et al.: Starcraft ii: A new challenge for reinforcement learning, *IJCAI-16* (2017).
[5] Terry, J. et al.: Pettingzoo: Gym for multi-agent reinforcement learning, *NeurIPS 2021*, Vol. 34, pp. 15032–15043 (2021).
[6] Juliani, A. et al.: Unity: A general platform for intelligent agents, *arXiv preprint arXiv:1809.02627* (2018).
[7] Cohen, A., Teng, E., Berges, V.-P., Dong, R.-P., Henry, H., Mattar, M., Zook, A. and Ganguly, S.: On the use and misuse of absorbing states in multi-agent reinforcement learning, *arXiv preprint arXiv:2111.05992* (2021).

ⓒ2023 Information Processing Society of Japan