深層強化学習モデルの内部挙動の言語化を通じた制御手法の構築

圓田彩乃　小林一郎
お茶の水女子大学
{g1720506,koba}@is.ocha.ac.jp

概要

本研究は説明可能 AI のひとつのアプローチとして，入力が画像である制御タスクに対し学習済みモデルの入出力関係から制御規則を生成し，これを言語化することで学習済みモデルの内部挙動を説明することを目指す．言語化された制御規則を用いて対象が制御可能であることを示すとともに，言語で記述された制御モデル内部の振る舞いに対する解釈可能性の向上を実現した．制御対象として， Space invaders を取り上げ実験を行なった結果，本研究の手法で構築された制御器は，学習済みモデルと同等のレベルの frame 数までプレイヤーが生存し，報酬は学習済みモデルと比較すると低いものの，複数の敵を倒して報酬の獲得に成功した．

1　はじめに

近年，様々な場面で深層学習が用いられている．深層学習は人間よりも高い精度で認識や予測を行うことがある一方で，深層学習モデルの内部挙動はブラックボックスであることから用途によっては使用が制限される．そのため，構築したモデルの内部挙動を捉える手法として，説明可能 AI の研究が盛んになっている．説明可能 AI は高いレベルの精度を保ちつつ， より説明可能なモデルを作り出す機械学習手法の構築を目指している [1, 2]．また，機械学習手法に対する説明性の評価なども検討されている [3]．説明手段のひとつとして，ニューラルネットワークからルールを抽出する DeepRED [4] などもある．そのような説明可能 AI のひとつのアプローチとして，本研究では深層学習モデルの内部挙動が人間が理解できるように言葉で説明することを目指す．アプローチ方法として，深層学習モデルで得られた入出力関係をファジィモデリング [5] し，その関係をファジィ言語変数からなる規則で表現することにより，モデルの入出力の振る舞いを人間が把握しやすいようにする．

本研究では入力が画像情報である Atari の Space Invaders[1)] を制御タスクとして設定し，言語化された規則を用いて制御を行い，その制御精度を確認する．また，制御規則をそのまま出力するのではなく，解釈可能性が向上するように制御規則の要約も併せて行う．

2　関連研究

Greydanu らの研究である Visualize Atari [6] は， 6 種類の Atari のゲームを制御タスクとして設定し，Saliency [7] をヒートマップとして入力画像に重ねることでモデルで学習したエージェントの内部挙動を可視化する説明 AI モデルを構築した．Saliency とは，学習済みモデルが画像のどの部分を注視して予測を行っているのかを可視化した手法である．この研究では学習済み深層強化学習モデルとして Baby A3C [8] を使用している．これは Asynchronous Advantage Actor-Critic (A3C) [9] モデルをカスタムしやすいようにコンパクトにしたモデルである．

Saliency Map の可視化を行うことで，モデルがどのような戦略を学習して実行しているのかだけでなく，学習過程でどのように戦略が進化しているか，そして制御を失敗してしまう原因となっている戦略や行動を分析することに成功している．その一方で，説明は入力画像上に注視箇所をヒートマップとして重ねて可視化することにとどまっており，説明文は画像を見ながら人手で作成している．

本研究では，学習済みモデルが制御の際に注視している個所を取り出す手法として Visualize Atari を使用し，制御中の学習済みモデルの内部挙動を自動的に言語で説明することに試みた．

1） https://www.gymlibrary.dev/environments/atari/space_invaders/

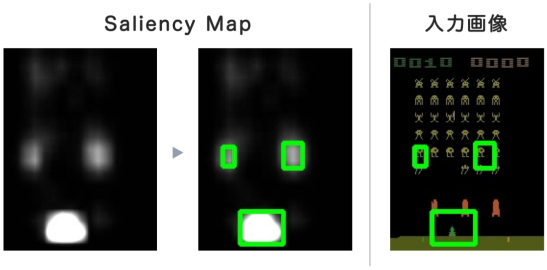


図 1　Saliency Map から Saliency 箇所を取り出した例

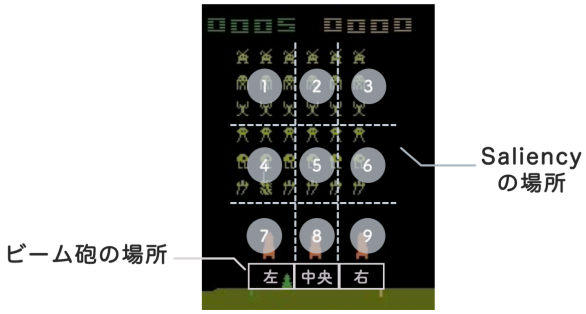


図 2　ビーム砲の場所と Saliency の場所の分割

3　提案手法

学習済み深層学習モデルの内部挙動を入出力情報をもとに獲得し，それを言語化して制御実験に使用する手法を説明する．

3.1　制御規則作成方法

入出力関係の獲得　Visualize Atari に用意されている学習済みモデルを使用し，入出力情報と Saliecny Map を獲得する．

Saliency 箇所の取り出し　Saliency Map をグレースケール化し，Open CV の輪郭抽出・外接矩形の関数を使用して入力画像の Saliency がかかっている場所を長方形に取り出す．（図 1 参照）

Saliency 内容の分類　取り出した Saliency 箇所に写っているものを分類する．分類モデルは，事前学習済み画像分類モデルを 2 エピソード分の切り取り済み Saliency 画像を用いてファインチューニングして作成する．

分類ラベルはファインチューニングに使用する切り取り済み Saliency 画像に写っているものを元に作成し，それぞれの画像には人手で分類ラベルを付与した．また，図 1 の右上の画像のように写っているものが切れてしまっている場合，それが何であるかを視認できる場合は分類ラベルを付与した．

制御規則生成　獲得した学習済みモデルの入出力情報と Saliency Map を用いて制御規則を生成する．

制御規則は以下の前件部・後件部のテンプレートを穴埋めする形で作成する．

> ビーム砲が *** (ビーム砲の場所... 左/中央/右) にいて，*** (Saliency の場所...①~⑨) に *** (Saliency の分類ラベル) があるとき，Action *** をとる．

図 3　制御器の構造 概要

エージェントが操作するビーム砲と Saliency の場所は図 2 のようにそれぞれ分割し，ビーム砲は中心の座標，Saliency 箇所は切り出した長方形の中心の座標からどの空間に属するかを算出する．

図 1 のように一つの入力画像に 2 箇所以上 Saliency がかかっていた場合，「ビーム砲が**左**にいて，④に**インベーダー単数**がある かつ ⑥に**インベーダー複数**がある かつ ... があるとき，Action **FIRE** をとる．」のようにテンプレートの前件部を AND で繋いで一つの制御規則とする．

3.2　制御器の構造

制御器の構造の概要を図 3 に示す．制御器には Space Invaders の画面が入力として与えられる．制御器は 3.1 で獲得したデータのうち，最も報酬が高いエピソードの Saliency Map からその frame において見るべき場所を判断し，その場所に何が映っているかを 3.1 で作成した分類モデルで分類する．併せて，入力画像を見てビーム砲の場所と Saliency がかかっている箇所を取り出す．それらの情報を組み合わせて該当する制御規則を抽出し，その個数が最も多い Action を実行する．

4　制御規則の言語化

実験では 1 frame ごとに Action を決定するのに使用した制御規則を言語化することで，学習済みモデ

ルの内部挙動の説明を行う．

3.1 で生成した制御規則は，学習済みモデルがどのような入力画像が与えられたときにどの Action をとったのかを捉えて表現することができている．しかし，1 frame ごとに使用した制御規則をそのまま言語化して出力すると，解釈可能性の低い説明文となってしまう．

そこで本研究では，生成された制御規則を要約して言語化することを試みる．具体的には，以下のテンプレートのように，制御規則ではそのまま記述しているビーム砲と Saliency の場所を，それらの位置関係（右上の高いところ，など）で記述することによって制御規則を要約を行う．

> ビーム砲の *** (ビーム砲と Saliency の場所の位置関係... 左/右/真上の高い/低い) ところに *** (Saliency の分類ラベル) があるとき，Action *** をとる．

x 軸方向の関係をビーム砲の場所と Saliency の長方形の中心の座標から「左上」「真上」「右上」の 3 通り，y 軸方向は Saliency の長方形の中心の座標から「高い」「低い」の 2 通りで表現し，これらを組み合わせることでビーム砲と Saliency 箇所の位置関係を表す．このように変更することで，変更前はビーム砲と Saliency 場所がそれぞれの分割数の組み合わせである 27 通りの表現があるのに対し，変更後は 6 通りの表現で制御規則を要約することができる．

制御規則の要約を行うことで解釈可能性が向上するだけでなく，異なる状況においても同じ制御規則が適用されるため，言葉を再利用することができるようになる．

5 実験

本研究では 3.1 の手法で生成した制御規則を使用した制御実験（実験 1）と 4 の手法で要約した制御規則を使用した制御実験（実験 2）を行った．

5.1 実験設定

入出力関係と Saliency Map の獲得 Visualize Atari における Space Invaders の control agent を使用し，学習済みモデルの入出力関係と Saliency Map を 50 エピソード・56,850 frame 分獲得した．なお今回の実験では，獲得した報酬が Baby A3C モデルの平均報酬である 550 を超えているエピソードのみを抽出して入出力関係などを獲得している．

表 1 分類ラベル

1 種類	ビーム砲，安全地帯，背景，残機， インベーダー単数，インベーダー複数
2 種類	ビーム砲+ビーム ビーム砲+安全地帯 ビーム+安全地帯 ビーム+インベーダー複数 安全地帯+インベーダー複数
3 種類	ビーム砲+ビーム+安全地帯 ビーム砲+安全地帯+インベーダー複数

Saliency 内容の分類モデルの作成 学習済みモデルから獲得した 50 エピソード分の入出力データと Saliency Map のうち 2 エピソード分のデータを使用し，切り出した Saliency がかかっている箇所に写っている内容を分類するモデルを作成する．

2 エピソード分の切り出した画像を目視で確認し，用意した分類ラベルを表 1 に示す．

Saliency 箇所を切り出した画像と分類ラベルを使用して，Torch Vision の ResNet のファインチューニングを行い，分類精度が 90% 近くになった．

制御規則 3.1 の手法で 9,241 個の制御規則が生成された．これに 4 の手法を用いると，制御規則が 6,853 個に要約された．

5.2 実験結果

実験 1・2 において，制御実験をそれぞれ 100 エピソード行った．

実験 1・2 と入出力関係などを獲得した学習済みモデルの実験時における獲得した報酬と生存 frame 数の結果をそれぞれ表 2 と表 3 に示す．

実験 1・2 と学習モデルの制御精度を比較すると，実験 1・2 のどちらも平均生存 frame 数は学習済みモデルと同等のレベルに達しており，平均報酬は学習済みモデルと比較すると低いものの，複数のインベーダーを倒して報酬を獲得することに成功した．

制御規則を要約していない実験 1 と要約している実験 2 の制御精度を比較すると，報酬・生存 frame 数ともに平均値に大きな違いはないが，実験 1 と比較すると実験 2 は報酬・生存 frame 数の最大値と最小値の差が大きく，不安定な制御となっていることがわかった．

実験時には，1 frame ごとに Action を決定するために使用した制御規則の言語化を行い，その Action

表 2　実験結果 - 報酬

	平均報酬	最大報酬	最小報酬
実験 1	223.85	390	180
実験 2	232.35	750	35
Baby-A3C	618	785	550

表 3　実験結果 - 生存 frame 数

	平均 frame	最大 frame	最小 frame
実験 1	938.26	1491	837
実験 2	919.62	1919	367
Baby-A3C	1137.7	1600	743

が選択された理由を説明した．例えば実験 2 において，図 4 の入力画像に対し「ビーム砲の **左上の低い** ところに **インベーダー単数** があるとき，Action **LEFT** をとる.」の制御規則が言語化された．

実験 2 における言語化された制御規則の他の例を付録で紹介している．

5.3　考察

実験 1・2 と学習済みモデルの比較　生存 frame 数は同等のレベルであるのに報酬に差が出てしまっている原因として，実験 1・2 は攻撃をしない Action が選択されることが多いことや，インベーダーからの攻撃から逃げるためにインベーダーの真下にいない frame が多いことが考えられる．また，制御規則を生成するにあたり，学習済みモデルがとっていたと思われる，より報酬の高い敵を優先して倒す戦略を踏襲できていないことも原因となっている可能性がある．

制御規則の要約あり/なしの比較　報酬と生存 frame 数の平均値をみると，実験 1・2 は同等の制御精度と言える．一方で，実験 2 において，報酬と生存 frame 数の最小値が実験 1 と比較してとても小さくなっている．これは，制御規則を要約したことによりビームを避けたり安全地帯の下に逃げるような制御規則が減ったことが原因として考えられる．制御規則は内容と個数が多いほどより多くの状況に対応できるが，一方で説明文としては解釈可能性が低くなってしまう．したがって，制御精度と説明文の解釈可能性のバランスを取ることが重要である．

制御の様子　制御の様子を観察したところ，実験 1・2 で作成した制御器はインベーダーの位置に関係なく画面左側に滞在し，インベーダーがビーム砲の真上に来るのを待ち構えて攻撃していた．一方，学習済みモデルはインベーダーの動きに合わせてビーム砲を移動しながら攻撃しており，構築した制御器は学習済みモデルの挙動を踏襲しきれていないことがわかった．これは複数の制御規則が該当した際の Action の決定方法や制御規則の作成方法などが原因として考えられる．制御器がより学習済みモデルの挙動を模倣できるよう，手法の改良を行う必要がある．

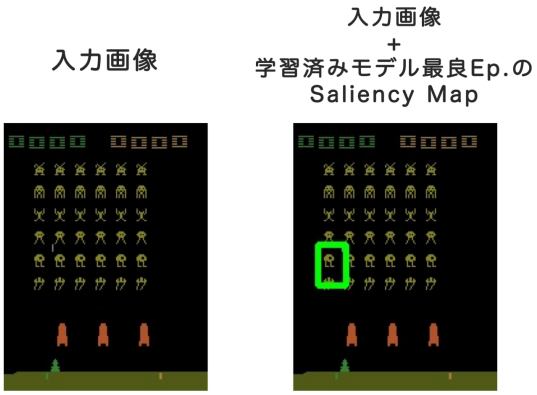


図 4　入力画像の例

6　おわりに

本研究は，説明可能 AI のひとつのアプローチとして，学習済み深層強化学習モデルの入出力関係から制御規則を生成し，これを言語化することで学習済みモデルの内部挙動を説明することを試みた．併せて，説明された制御規則を用いて制御実験を行い，その制御精度を確認した．また，説明文の解釈可能性を向上させるため，生成した制御規則の要約も行った．

言語化された制御規則を用いて制御実験を行った結果，提案した手法で構築された制御器は，学習済みモデルと同等のレベルの frame 数まで生存することに成功した．報酬は学習済みモデルと比較すると低いものの，複数のインベーダーを倒して報酬を獲得することができた．また，説明文の解釈可能性を向上させるための要約をした制御規則を用いて制御実験を行った結果，要約する前の規則を使用した制御器と比較すると不安定な制御となるものの，概ね同等の制御精度で制御を行うことができた．

今後の課題として，制御器の制御精度をより向上させることと，ゲームに限らない入力情報が画像である様々な制御対象に対して使用できる制御器の構築手法となるように改良を進める．

謝辞

本研究の一部を JSPS 二国間共同研究（JPJSBP120213504）からの支援を受けました．ここに深謝いたします．

参考文献

[1] Sherin Mary Mathews. Explainable artificial intelligence applications in nlp, biomedical, and malware classification: A literature review. pp. 1269–1292, 2019.

[2] A. Adadi and M. Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). **IEEE Access**, Vol. 6, pp. 52138–52160, 2018.

[3] Leilani H. Gilpin, David Bau, Ben Z. Yuan, Ayesha Bajwa, Michael A. Specter, and Lalana Kagal. Explaining explanations: An approach to evaluating interpretability of machine learning. **CoRR**, Vol. abs/1806.00069, , 2018.

[4] Jan Ruben Zilke, Eneldo Loza Mencía, and Frederik Janssen. Deepred–rule extraction from deep neural networks. In **International Conference on Discovery Science**, pp. 457–473. Springer, 2016.

[5] 姜根澤, 菅野道夫. ファジィモデリング. 計測自動制御学会論文集, Vol. 23, No. 6, pp. 650–652, 1987.

[6] Samuel Greydanus, Anurag Koul, Jonathan Dodge, and Alan Fern. Visualizing and understanding atari agents. In **International conference on machine learning**, pp. 1792–1801. PMLR, 2018.

[7] Karen Simonyan, Andrea Vedaldi, and Andrew Zisserman. Deep inside convolutional networks: Visualising image classification models and saliency maps. **arXiv preprint arXiv:1312.6034**, 2013.

[8] Danijar Hafner, Timothy P Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models. In **International Conference on Learning Representations**, 2021.

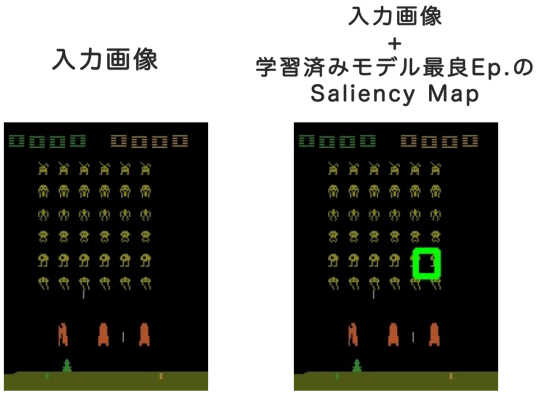
[9] Volodymyr Mnih, Adria Puigdomenech Badia, Mehdi Mirza, Alex Graves, Timothy Lillicrap, Tim Harley, David Silver, and Koray Kavukcuoglu. Asynchronous methods for deep reinforcement learning. In **International conference on machine learning**, pp. 1928–1937. PMLR, 2016.

付録

ここでは，要約した制御規則を使用した制御実験時の制御の様子を紹介する．具体的には，ある frame における入力画像と，その frame において出力された「入力画像に対してどの Action を決定したのか」を説明する文章を例示する．

例 1：

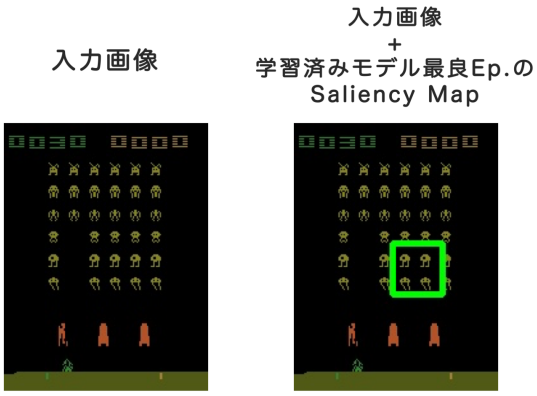
入力画像：



言語化された制御規則：
ビーム砲の **右上の高い** ところに **インベーダー単数** があるとき，Action **RIGHT FIRE** をとる．

例 2：

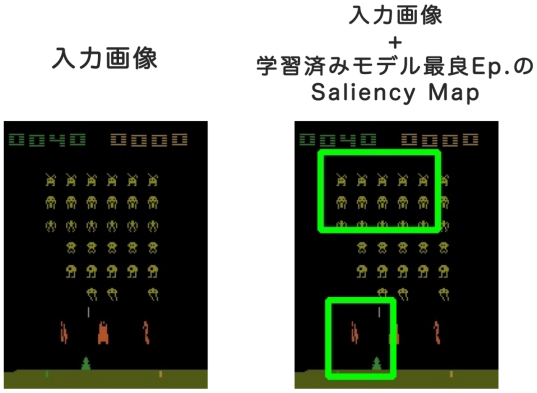
入力画像：



言語化された制御規則：
ビーム砲の **右上の低い** のところに **インベーダー複数** があるとき，Action **RIGHT FIRE** をとる．

例 3：

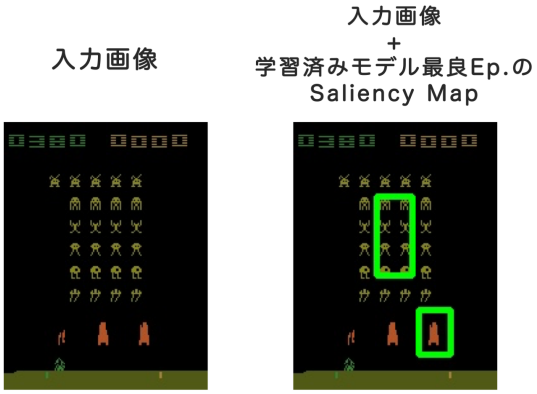
入力画像：



言語化された制御規則：
ビーム砲の **右上の高いところ** のところに **インベーダー複数** がある かつ **真上の低い** のところに **ビーム砲と安全地帯** があるとき，Action **RIGHT FIRE** をとる．

例 4：

入力画像：



言語化された制御規則：
ビーム砲の **右上の高い** のところに **インベーダー複数** がある かつ **右上の低い** のところに **安全地帯** があるとき，Action **FIRE** をとる．
