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

理学専攻・情報科学コース　圓田 彩乃　(指導教員：小林 一郎)

1　はじめに

近年様々な場面で活用されている深層学習は内部挙動がブラックボックスであることが問題視されている．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．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．本研究では入力が数値情報である CartPole と，入力が画像情報である Atari の Space Invaders の 2 種類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．

2　入力情報が数値である制御器構築

2.1　制御器構築方法

ファジィ制御は，入出力の関係をファジィ集合を用いた制御規則で表現し，制御規則の表現自体に曖昧さを含むことを許容することにより，数式で表現しにくい制御対象などにも頑健な制御を実現可能とする手法である [1]．ファジィ制御器は，与えられた入力を制御規則における前件部のメンバーシップ関数で評価され，その適用度を後件部に伝え，複数の制御規則の後件部の値の重みつき和を出力する．本研究では Deep Q-Network (DQN) [2] で構築された制御器から得られた入出力関係からファジィ制御規則を導き出すことで，DQN の内部挙動を捉え，ファジィ制御規則を言語モデリングすることでその挙動を言語で説明可能とする．

本研究のファジィ制御器構築方法の概要を図 1 に示す．

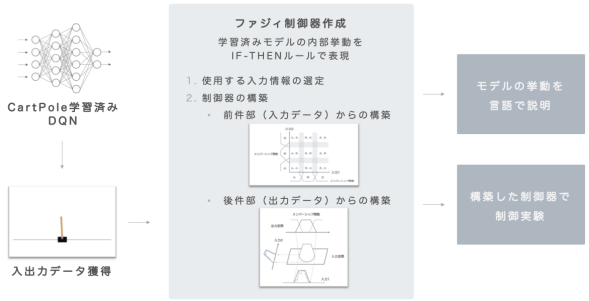

図 1: CartPole - 制御器構築方法 概要

本研究では，入力データから制御規則生成を「前件部からのファジィ制御器構築」，出力データから制御規則生成を「後件部からのファジィ制御器構築」とし，それぞれの方法でファジィ制御器構築を試みる．いずれの制御器構築手法でも，制御則のファジィ集合表現には台形型のメンバーシップ関数を使用する．

2.2　実験結果

本研究では，制御器が取れる Action を「1 ～ 20 の力で左/右に押すことと押さないことの 41 種類」(実験 1) と「10 の力で左/右に押すことと押さないことの 3 種類」(実験 2) とした 2 種類の実験設定で実験を行った．

実験 1：Action が 41 種類の場合　DQN は 342 エピソードで学習が完了し，そのあと 200step 倒立できた 100 エピソード分の入出力データを獲得した．

制御規則は出力データを獲得したエピソードで実際に行われた Action 数と同じ 35 個生成された．

構築したファジィ制御器で CartPole を 100 エピソード制御した結果，倒立平均 step 数は 162.3 step，100 エピソードの中で 200 step 到達できたのは 38 エピソードであった．100 エピソード中 89 エピソードで 100 step 到達していることから DQN 学習後のモデルの精度とまではいかないまでも CartPole をある程度制御できていると言える．

実験 2：Action が 3 種類の場合　DQN は 421 エピソードで学習が完了し，そのあと 200step 倒立できた 10000 エピソード分の入出力データを獲得した．

この実験では，前件部から/後件部からそれぞれファジィ制御器を構築し，制御実験を行った．前件部からのファジィ制御器構築では，制御器の構造や構築に使用するデータを変えて 4 つの実験を行った．この実験では，1 エピソードを「立たせる挙動」「中間の挙動」「キープの挙動」のフェーズに分けてそれぞれでファジィ制御器を作成し，経過 step 数からこれらのファジィ制御器を使い分けることで，1 エピソード中における振る舞いの変化に対応できるようにした．この実験設定を表 1 に示す．

表 1: CartPole - 実験 2 実験設定

実験	フェーズ	使用データ
実験 2-1	立たせる挙動	1 ～ 11 step
実験 2-2	立たせる挙動	1 ～ 20 step
	キープの挙動	150 ～ 200 step
実験 2-3	立たせる挙動	1 ～ 125 step
	キープの挙動	150 ～ 200 step
実験 2-4	立たせる挙動	1 ～ 50 step
	中間の挙動	50 ～ 125 step
	キープの挙動	150 ～ 200 step

後件部からのファジィ制御器構築では，前件部から構築したファジィ制御器の中で最も制御精度が良かった実験 2-4 を参考に，同じデータを使用して「立たせる挙動」「中間の挙動」「キープの挙動」でそれぞれファジィ制御器を構築した．

各実験において作成されたファジィ制御器で，CartPole を 10 エピソード制御した結果を表 2 に示す．

表 2: CartPole 制御結果

実験	倒立平均 step 数	制御規則数
前件部 - 実験 2-1	119.7	173
前件部 - 実験 2-2	150.6	226/204
前件部 - 実験 2-3	160.6	329/204
前件部 - 実験 2-4	171.9	291/239/204
後件部から構築	120.7	3/3/3
学習済み DQN	200.0	-

2.3 考察

それぞれの実験での制御性能と作成されたファジィ制御器の言語規則数と比較した結果,「ファジィ制御器における CartPole の制御性能」と「生成された制御規則の個数＝分かりやすさ」はトレードオフの関係になっていると考えられる.

3 入力情報が画像である制御器構築

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，説明可能 AI の先行研究である Visualize Atari [3] を使用し，制御中の学習済みモデルの内部挙動を言語で説明することに試みた.

3.1 制御器構築方法

制御器構築方法の概要を図 2 に示す.

図 2: Space Invaders - 制御器構築方法 概要

3.2 制御器の構造

制御器は獲得した学習済みモデルのデータのうち，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，そこに何が映っているかを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．これらの情報から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.

3.3 制御規則の削除

生成され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ため，制御規則を削除して言語化することを試みる．具体的には，制御規則では絶対的な位置で記述しているビーム砲と Saliency の場所をそれらの相対的な位置（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削除し，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内部挙動の説明を行う.

制御規則の削除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，言葉を再利用できるようになる.

3.4 実験結果

実験 1 では生成された制御規則を，実験 2 では要約した制御規則を使用して制御器を構築し，それぞれ 100 エピソード実験を行った．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 数の結果を表 3 に示す.

表 3: Space Invaders 制御結果

	報酬			生存 frame 数		
	平均	最大	最小	平均	最大	最小
実験 1	224	390	180	938	1491	837
実験 2	232	750	35	920	1919	367
Baby-A3C	618	785	550	1138	1600	743

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った．例えば実験 2 において,「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.

3.5 考察

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，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．これは，制御規則を要約したことによりビームを避けるような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.

4 おわりに

本研究では深層学習モデルの内部挙動を人間が理解できるように言葉で説明することを目指し，学習済みモデルの内部挙動を模した制御器を構築し，言語化した規則を使用して制御実験を行った．実験の結果，学習済みモデルの精度まではいかないまでもある程度の精度で制御を行い，報酬を獲得することに成功し，制御精度と説明文の解釈可能性がトレードオフの関係になっていることが分かった.

今後の課題として，トレードオフを軽減することと，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.

参考文献

[1] 菅野道夫. ファジィ制御. 日刊工業新聞社, 1988.

[2] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *nature*, Vol. 518, No. 7540, pp. 529–533, 2015.

[3] 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.