ロバスト IoT×AI のための深層防御システムの研究調査

研究代表者　　川　本　一　彦　　千葉大学大学院情報学研究院
共同研究者　　計　良　宥　志　　千葉大学大学院情報学研究院

1 序論

近年、深層学習および深層強化学習の飛躍的な進展により、画像認識からロボット制御に至るまで、さまざまなタスクで高性能な AI システムの構築が可能となってきた。特に IoT（Internet of Things）技術と深層学習を融合した IoT × AI システムは、Society 5.0 が目指す超スマート社会の中核的な情報処理基盤として、産業界および社会から大きな注目を集めている[1][2]。しかしながら、これらのシステムは、実際の使用環境において外乱や予期せぬ環境変動にさらされるだけでなく、AI システムの誤作動を引き起こす目的で意図的に設計された攻撃（敵対的攻撃）にも直面するリスクがあり、安全性や信頼性を確保する上で重大な課題を抱えている[3][4]。このため、システムのロバスト性（さまざまな外部摂動に対する耐性）の向上が、IoT × AI システムの社会実装を推進する上で極めて重要である。

本研究では、敵対的強化学習というアプローチを軸にして、深層学習モデルのロバスト性を体系的に評価し、その向上手法を明らかにすることを目的としている。敵対的強化学習とは、モデルを学習する際に、故意に作成された悪意のある入力（敵対的摂動を含んだデータ[5]など）を訓練データとして用いることで、AI モデルが攻撃やノイズに強くなるように訓練する手法である。意図的に摂動を加えたデータを追加することでデータの多様性を高め、モデルの頑健性を向上させる。本研究では、この手法を活用し、モデルのロバスト性を実用的に高めるための技術を検討する。

具体的な対象領域として、本研究は二つのタスクを設定した。一つ目は、オンラインで学習する深層強化学習を用いた歩行ロボット制御であり、二つ目は、あらかじめ収集された固定のデータセットだけを使って学習するオフライン強化学習を用いたロボット制御である。これらのタスクにおいて、ロボットの身体や制御信号に意図的な摂動（敵対的摂動）を導入し、その状況下でも安定した動作ができるように、敵対的強化学習を実施する。さらに、オフライン強化学習で得られた方策をオンライン環境で微調整することで実際の環境への適応力を高める、オフライン-オンライン強化学習についても検討する。

従来、画像認識を中心に敵対的訓練の有効性が研究されてきたが[4]、ロボット制御のような動的な制御タスクで、状態や行動に対する敵対的摂動を同時に考慮した研究例は限定的であった。特にオフライン強化学習では、固定データを用いる特性から、未知の摂動に対して脆弱であることが課題として挙げられている[6-9]。そこで本研究では、オンライン強化学習、オフライン強化学習、そして両者を統合したオフライン-オンライン手法のそれぞれについて、動的制御タスクにおけるロバスト性向上のための包括的なアプローチを提示し、その有効性を検証する。

2 関連研究

深層学習や強化学習を用いた AI システムの実環境適用において、外部摂動に対する脆弱性は古くから知られており、これに対処するための研究が活発に行われている。特に、敵対的攻撃への対策、ドメインランダム化による汎化能力の向上、そしてデータ拡張による頑健性強化などが、近年の主要な研究トピックとなっている。

2-1 敵対的攻撃と敵対的訓練

深層学習における敵対的攻撃（adversarial attack）は、入力データに対して人間には知覚困難な摂動を加えることで、モデルの予測を誤らせる攻撃手法であり、画像分類や物体認識の分野で広く研究されてきた[10]。代表的な攻撃手法として Projected Gradient Descent（PGD）[5] が知られている。これに対する代表的な防御手法として、敵対的摂動をあらかじめ訓練データに含める敵対的訓練（adversarial training）があり、一定のロバスト性向上効果が確認されている。また、ロボット領域では RARL[11]など、エージェン

トと敵対者を同時学習させて方策の頑健性を高める試みも報告されている。
我々はこの敵対的訓練の考え方を、歩行ロボットの制御タスクに拡張し、ロボットのアクチュエータ信号に対して意図的に摂動を加える攻撃に対する耐性を評価した。とくに、差分進化法を用いて敵対的トルク摂動を生成することで、ロボットのバランス崩壊を誘発する攻撃を再現し、それに対する防御の有効性を示した。

2-2 ドメインランダム化と環境変動への適応

シミュレーションと実環境との乖離、いわゆるリアリティギャップの克服は、ロボティクスにおける深層強化学習の重要課題である。これに対するアプローチの一つとして、ドメインランダム化（domain randomization）が提案されており、摩擦係数や地形の傾斜、体型パラメータなどをランダムに変化させた環境で訓練することで、ロバストな方策の獲得が可能となる[12]。さらに、モデル集約的に最悪ケースを選ぶ EPOpt[13] など、頑健性を目的とした探索的ドメインランダム化も提案されている。
我々の研究では、このドメインランダム化と敵対的訓練を比較し、敵対的トルクによる訓練の方が、ランダムノイズに対しても高い汎化性能を示すことを明らかにした。これは、摂動の構造的特性を利用することで、より本質的なロバスト性を実現できる可能性を示唆している。

2-3 オフライン強化学習におけるロバスト性

近年注目を集めているオフライン強化学習は、事前に収集されたデータセットのみを用いて方策を学習する手法であり、データ収集コストや安全性の観点からロボティクス分野への応用が期待されている[14]。しかし、訓練中に探索を行わないという特性上、未知の環境や摂動への適応力は限定的であり、そのロバスト性は重大な課題である。
既存研究では、価値推定の安定化や分布制約による Q 値の過大推定を抑えることが主な焦点とされてきた。たとえば CQL[7] は行動価値の過大推定を抑制することで分布外行動を制限し、テスト時の性能劣化を軽減することを示している。また BCQ[6]、TD3+BC[8]、IQL[9]といった代表的手法も提案されているが、アクチュエータ信号など行動空間への摂動に対するロバスト性については十分に検証されてこなかった。
本研究では、BCQ・TD3+BC・IQL それぞれに対してランダムおよび敵対的摂動を加えた実験を行い、その脆弱性を定量的に明らかにした。さらに、従来のオフライン学習にオンラインでの少数エピソードを追加するオフライン-オンライン学習を提案し、行動摂動に対するロバスト性を大幅に向上できることを示した。

3 手法

本節では、本研究で提案・採用したロバストな強化学習のための手法について説明する。具体的には、オンライン強化学習とオフライン強化学習を対象とし、敵対的強化学習を中心にその手法を述べる。敵対的強化学習とは、あらかじめ学習済みのモデルを用いて強力な敵対的サンプルを生成し、それらを通常のデータセットに追加することで、多様かつ攻撃に対して耐性の高いデータセットを構築する手法である。この手法では、敵対的サンプルを事前に作成し、通常のデータと併せて再訓練する。一方、敵対的サンプルを訓練途中で逐次生成して学習する敵対的訓練もあるが、学習途中で攻撃データを新たに生成する必要がないため、計算コストや実装の効率性において利点がある。
さらに、本研究ではオフラインで学習したモデルをオンライン環境に転移させるために、オフライン-オンライン強化学習の枠組みを採用している。ここでは、オフライン環境で獲得した方策を、敵対的データを活用したオンライン環境でさらにファインチューニングすることで、実環境や未知の摂動に対するロバスト性の向上を狙っている。この敵対的ファインチューニングによって、オフラインで学習した方策がオンラインでの新規の敵対的攻撃にも迅速に適応可能となることが期待できる。

3-1 敵対的オンライン強化学習

オンライン深層強化学習におけるロバスト性向上のために、敵対的強化学習を採用する。この手法では、学習過程において意図的に設計された摂動を導入することで、エージェントが最悪条件に直面しながら方策

を学習することを促す（図１）。

具体的には、ロボット制御タスクを対象に、差分進化アルゴリズム［17］を用いてエージェントの報酬を最小化するような敵対的摂動を生成し、ロボットの制御信号（関節トルク指令値）に加えることで学習を困難化させる。このような敵対的強化学習の一手法として敵対的データ拡張（Adversarial Data Augmentation）がある。これは、あらかじめ学習済みのモデルを用いて敵対的な摂動を加えたデータを生成し、それらを通常の訓練データに追加して再学習することで、攻撃耐性のある方策を得る方法である。また、もう一つの手法である敵対的ファインチューニング（Adversarial Fine-Tuning）では、通常環境で訓練済みのモデルに対して、敵対的な摂動を含む環境で短期間の追加訓練を行う。このアプローチにより、既存の方策をベースに、未知の外乱や敵対的攻撃に対して迅速に適応可能なロバストな方策の獲得を目指す。

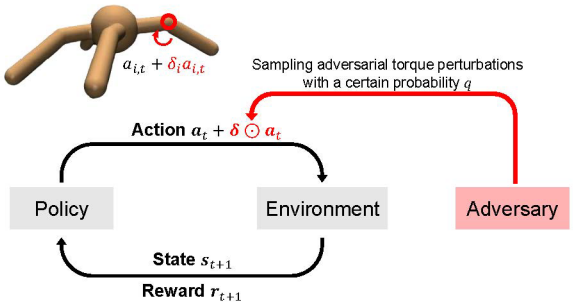

図 1 オンライン強化学習に対する敵対的強化学習

3-2 敵対的オフライン強化学習

オフライン強化学習では、環境との追加的なインタラクションなしに、あらかじめ収集されたデータのみを用いて方策を学習する。そのため、学習時のデータと実際の環境との分布のズレが大きな課題となる。これを抑制するため、方策がデータの分布から大きく逸脱しないよう、制約や正則化を導入する方法が用いられている。しかしながら、こうした保守的な学習戦略は、実環境での予期せぬ状況に対して脆弱性を招く可能性がある。

本研究では、このような脆弱性を明確に捉えるため、学習済み方策に対して関節トルクに基づく摂動を加える３つの環境（通常、ランダム、敵対的）を設定した。特に敵対的摂動は、差分進化アルゴリズムにより最適化されたものであり、エージェントの報酬を最も低下させるよう設計されている。さらに、図２に示すように、オフライン強化学習における敵対的訓練の枠組みを導入し、敵対的摂動下での方策の学習と評価を行う。複数のエピソードを通じてエージェントが転倒するか、または最大ステップ数に到達するまでの平均報酬を算出し、学習方針ごとのロバスト性を比較した。これにより、従来の保守的な学習アプローチが新たな脅威に対してどの程度有効かを検証した。

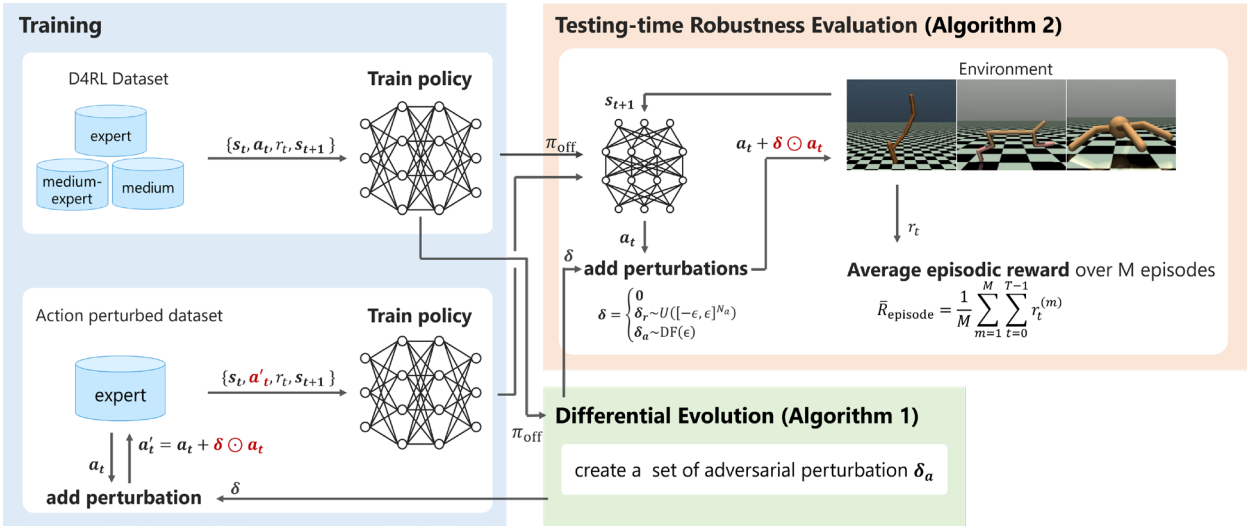

図 2　オフライン強化学習に対する敵対的強化学習

3-3 オフライン-オンライン強化学習

オフライン強化学習で獲得された方策は、固定されたデータセットの範囲内で学習されているため、実環境で新規の敵対的摂動や未知の外乱に直面すると性能が著しく低下することがある。この課題を解決するため、本研究ではオフライン-オンライン強化学習の枠組みを採用し、オフラインで事前学習したモデルに対してオンライン環境で短期間の追加学習（敵対的ファインチューニング）を行い、実環境へのロバストな転移を実現する。

具体的には、オフライン強化学習手法として TD3+BC を用いて事前学習した方策を初期方策として採用する。オンライン環境において敵対的摂動を導入した状態で少数のエピソードを用いたファインチューニングを実施することで、オフラインで得た方策を基盤に、実環境で発生しうる摂動に迅速かつ頑健に適応可能な方策を獲得することを目的とする。

4 実験

本節では、前節で説明した手法に基づき実施した一連の実験について述べる。具体的には、(1) 敵対的オンライン強化学習、(2) 敵対的オフライン強化学習、(3) 敵対的オフライン-オンライン強化学習の有効性の 3 点について、ロボット制御タスクを対象として評価した。実験では、Proximal Policy Optimization (PPO) アルゴリズム [18] を用いて強化学習をしている。

4-1 敵対的オンライン強化学習の評価

四足歩行ロボット Unitree A1 （図 3 左）を対象に、敵対的摂動に対するロバスト性を評価する。4.1 節では、関節トルクへの攻撃による影響、4.2 節では敵対的オンライン強化学習の有効性を評価する。

4-1-1 敵対的トルク信号に対するロバスト性評価

本実験では、MuJoCo シミュレータ [19] 上で四足歩行ロボット Unitree A1 に対する敵対的攻撃の影響を評価した。攻撃は関節トルクに対して差分進化により最適化された敵対的摂動を加える形で実施され、その強度は摂動パラメータ ε =0.0, 0.1, 0.2, 0.3, 0.4, 0.5 の 6 段階に設定された。各 ε に対して 1000 回のシミュレーションを行い、累積報酬の平均と分布を観測した。

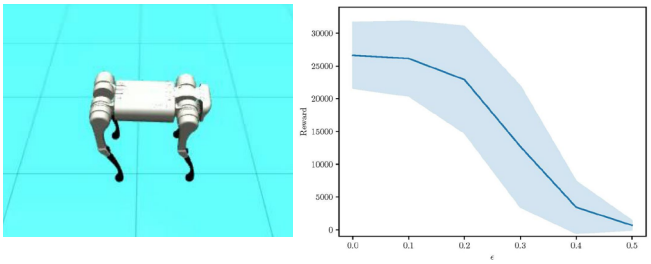

図 3　Unitree AI ロボット：概形（左）、敵対的トルク摂動に対する平均報酬分布

その結果、図 3 右に示すように、Unitree A1 における平均報酬は摂動強度の増加に対して減少し、ε =0.2 以上では性能劣化が急速に進み、ε =0.4 以上では敵対的トルク摂動が歩行タスクの実行に深刻な影響を及ぼすことが確認された。これにより、Unitree A1 は関節トルクに対して高い脆弱性を有しており、攻撃によってタスク遂行能力が著しく低下させることが可能であることが明らかとなった。

4-1-2 敵対的強化学習によるロバスト性向上評価

本実験では、四足歩行ロボット Unitree A1 を対象として、敵対的データ拡張（Adversarial Data Augmentation: ADA）および敵対的ファインチューニング（Adversarial Fine-Tuning: AFT）の有効性を検証し、それぞれの手法がもたらすロバスト性向上の違いについて検証する。

まず、ADA の訓練においては、摂動の強さを ε =0.5、摂動を導入する確率を q=0.01 に設定した。AFT では、通常環境であらかじめ訓練されたモデルに対して、学習率 3e-5、追加イテレーション数 50000、摂動確率 q=0.05 の条件で再訓練した。比較対象は、同じ ε および q の条件でランダムノイズを導入したドメインランダム化モデル（ランダム環境）である。

評価は、通常環境、ランダム環境、敵対的環境の 3 種類のテスト環境において実施し、それぞれの条件下で 1000 エピソードの歩行試行を行い、累積報酬の平均と標準偏差を算出した。表 1 に示されるように、AFT モデルはランダム環境において 26165、敵対的環境において 21148 と最も高い平均報酬を記録した。

表 1　敵対的強化学習によるロバスト性向上の評価

訓練/テスト	通常環境	ランダム環境	敵対的環境
通常環境	30445±6481	22890±10750	601±898
ランダム環境	**32375±6155**	23094±10886	1265±2122
ADA	19311±6238	15574±7354	2875±3773
AFT	28370±7146	**26165±8365**	**21148±11024**

これに対して、ADA モデルは通常環境で 19311、およびランダム環境で 15574、および敵対的環境で 2875 と全体的に性能低下が観察された。これは、訓練初期において敵対的摂動が過度に干渉したことで、右前足を用いない「三本足歩行」といった極端な戦略を学習してしまったことに起因すると考えられる。実際に観察された歩行動作では、明らかに非対称な運動パターンが確認されており、特定の攻撃に過剰適応した結果、通常環境における汎化性能が著しく損なわれていることが示唆された。

このような現象は、学習初期に敵対的な状態に遭遇した際、偶然にもその状況下で高い報酬が得られた場合に、エージェントがそれを好ましい方策と誤認し、その後の学習でその方策に収束してしまうことで発生する。つまり、早期の敵対的干渉が、探索のバイアスを生み出し、学習に悪影響を与える可能性がある。

以上の結果から、Unitree A1 におけるロバスト性向上には、敵対的ファインチューニングがより有効なアプローチであることが明らかとなった。AFT は、比較的少ない訓練ステップで高い適応性と安定性を実現できることから、実環境での応用においても有望な手法である。一方で、ADA は摂動に対する耐性について一定程度はあるものの、訓練の設計次第では過適応や戦略の偏りが生じるため、慎重な運用が求められる。

4-2 敵対的オフライン強化学習の評価

四足歩行ロボット Ant-v2（図 4）を対象に、オフライン強化学習（BCQ[6]、TD3+BC[8]、IQL[9]）のテスト時ロバスト性を評価する。評価は MuJoCo 物理シミュレータ上で実施され、特に関節トルク信号に対するランダムおよび敵対的摂動下での性能変化を測定する。オフライン用のデータセットは D4RL データセット[20]を用いる。オフライン手法とオンライン手法（PPO[18]）を比較し、さらに訓練データの構成やデータ拡張の効果についても検討する。

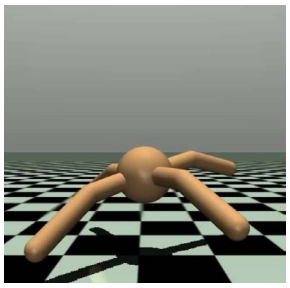

図 4 Ant ロボット

4-2-1 敵対的トルク信号に対するロバスト性評価

まず、Ant-v2 環境において、通常・敵対的の 2 種類の摂動条件下で各強化学習アルゴリズムの平均累積報酬を 1000 試行で評価した。表 2 に示すように、オフライン強化学習手法（BCQ, TD3+BC, IQL）は敵対的摂動条件下で著しい性能低下を示し、すべての手法で平均報酬が 0 を大きく下回った。具体的には、BCQ では −1129、TD3+BC では −1209、IQL では −1503 の平均報酬となっており、タスク遂行が困難な状態に陥っていることがわかる。一方、オンライン手法である PPO では、同条件下でも平均報酬が 2654 を維持しており、オフライン手法との間に大きな性能差が見られた。

表 2 オフライン強化学習とオンライン強化学習の比較

アルゴリズム	通常環境	敵対的環境
BCQ	4511±1568	−1129±1083
TD3+BC	4715±1676	−1209±1089
IQL	4175±1546	−1503±1026
PPO（オンライン）	5698±989	2654±1748

この結果は、オンライン強化学習が環境との相互作用を通じて多様な状態・行動の組み合わせを学習できるのに対し、オフライン強化学習は限られたデータに基づいて学習を行うため、予期しない外乱に対する適応力に欠けるという構造的な違いを反映している。とりわけ、敵対的な摂動が加わる状況では、環境への応答性や修正行動の獲得が重要であり、これがオンライン手法の優位性として現れているといえる。

4-2-2 敵対的強化学習によるロバスト性向上評価

行動空間に敵対的な摂動を加えたデータセットを用いた学習によって、オフライン強化学習モデルのロバスト性が向上するかを Ant 環境において検証した。しかし表３に示すように、明確な性能向上は確認されなかった。

表 3　敵対的オフライン強化学習によるロバスト性向上の評価

学習環境	アルゴリズム	通常環境	敵対的環境
通常環境	BCQ	4511±1568	−1129±1083
	TD3+BC	4715±1676	−1209±1089
	IQL	4175±1546	−1503±1026
敵対的環境	BCQ	1609±1685	621±1530
	TD3+BC	−159± 534	−428±815
	IQL	1758±1628	505±1414

まず､通常データで学習したモデルでは、通常環境における平均報酬は BCQ[7]で 4511、TD3+BC[8]で 4715、IQL[9]で 4175 と高い値を示しているが、敵対的環境下ではそれぞれ−1129、−1209、−1503 と大幅に低下している。これに対し、敵対的データで学習したモデルでは、通常環境においてもすでに BCQ が 1609、TD3+BC が−159、IQL が 1758 と大きく性能が下がっており、さらに敵対的環境ではそれぞれ 621、−428、505 と著しい劣化が見られる。

特に TD3+BC においては、通常環境での平均報酬が負の値−159 を示しており、基本的な歩行動作すら成立していない可能性がある。このことから、敵対的摂動を含むデータをそのまま学習に用いるだけでは、ロバスト性の向上どころか、通常環境における基本性能をも損なうリスクが高いことが明らかとなった。

以上の結果は、Ant ロボットのような複雑な制御タスクにおいて、静的データのみに依存したオフライン学習では、敵対的環境への適応が困難であることを示唆している。したがって、オフラインで学習したモデルを初期モデルとし､少量のオンライン相互作用により方策を適応させる｢オフライン−オンライン強化学習｣などのハイブリッド手法の導入が、ロバスト性の獲得に有効であると考えられる。

4-3 オフライン-オンライン強化学習における敵対的ファインチューニングの評価

本実験では、オフライン強化学習済みモデルをオンライン環境でファインチューニングすることで、通常および敵対的環境における Ant-v2 制御性能がどのように向上するかを検証した。

表 4　オフライン−オンライン強化学習によるロバスト性向上の評価

学習方法	学習環境	通常環境	敵対的環境
オフライン	通常環境	4715±1676	−1209±1089
オフライン−オンライン	通常環境	5005±1035	−286±1500
	敵対的環境	5106±1146	3099±1165

オフライン強化学習として TD3+BC[8]を用い、500 万ステップ学習したモデルを初期モデルとした。データセットは D4RL[20]を用いた。オフライン−オンライン強化学習のファインチューニングでは、オンライン環境で 100 万ステップ行った。敵対的な摂動強度は $\varepsilon=0.5$ とした。

表４に実験結果を示す。結果から、オフライン−オンライン強化学習は、通常環境および敵対的環境のいずれにおいても性能を向上させた。特に、敵対的ファインチューニングによって、敵対的環境下での性能が大幅に改善され、モデルの頑健性が強化されたことが確認された。具体的には、オフライン強化学習のみのモデルでは、敵対的環境において平均報酬が−1209 と大きく低下していた。これに対し、オフライン−オンライン強化学習により敵対的環境下での平均報酬は 3099 まで大幅に向上した。この結果は、オンライン環境で敵対的摂動を考慮したファインチューニングを行うことで、モデルが未知の外乱にも適応可能となったことを示している。また、通常環境においても一定の性能向上が見られ、敵対的ファインチューニングがモデル全体の性能向上に寄与していることが確認された。

５結論と今後の課題

本研究では、深層強化学習を用いたロボット制御タスクにおける AI モデルのロバスト性向上を目指し、オンライン強化学習、オフライン強化学習、およびオフライン−オンライン強化学習という三つのアプローチを評価した。特に、敵対的データ拡張や敵対的ファインチューニングといった手法が、未知の環境や外乱に対して AI の適応能力を高めるうえで有効であることを示した。

オンライン強化学習では、敵対的ファインチューニングにより、環境変動や外乱下でも安定した動作が可能な方策を獲得できた。一方、オフライン強化学習は静的なデータに依存するため、未知の摂動に対してロバスト性に限界があることが確認された。これに対し、本研究で採用したオフライン-オンライン強化学習の枠組みでは、オフライン学習済みの方策をオンライン環境で効率的に適応させ、外乱環境でも性能を維持できることを確認した。

これらの成果は、今後のロバストな IoT×AI システムの設計において有用な知見となると考えられる。多くの IoT デバイスは、センサやアクチュエータを通じて物理世界と密接に連携しており、環境変動や外乱に常にさらされている。そのため、エッジ AI や IoT ロボティクス領域で高信頼性の AI 制御を実現するには、AI モデルが環境の多様な変化に頑健に対応できる設計が不可欠である。本研究で検討した手法は、こうしたリアルタイム性や安全性が求められる IoT 応用分野に向けた基盤技術として大きな可能性を持つ。

今後の課題としては、オンライン適応のさらなる効率化とデータコストの低減が挙げられる。IoT 機器は計算資源や通信帯域が限られる場面も多いため、軽量かつ高効率なロバスト学習手法の開発が求められる。また、ファインチューニング時の摂動設計や自動適応機構の高度化を進め、より広範な環境変化に対する汎化性能を高めることが必要である。さらに、今回の成果はシミュレーション環境での評価に基づくものであり、次の段階では物理ロボットや実際の IoT デバイスへの適用と実証が必要である。特に、IoT を活用したスマートファクトリーやスマートシティといった領域では、現実の運用環境での AI の頑健性が重要であり、そのための信頼性評価や説明可能性（explainability）の向上も必要である。

【参考文献】

[1] J. Gubbi, R. Buyya, S. Marusic, M. Palaniswami, Internet of Things (IoT): A Vision, Architectural Elements, and Future Directions, Future Generation Computer Systems, vol.29, no.7, pp.1645-1660, 2013.

[2] T. J. Saleem, M. A. Chishti, Deep Learning for Internet of Things Data Analytics, Procedia Computer Science, Vol.163, pp.381-390, 2019.

[3] C. Szegedy, W. Zaremba, I. Sutskever, Intriguing Properties of Neural Networks, ICLR, 2014.

[4] I. J. Goodfellow, J. Shlens, C. Szegedy, Explaining and Harnessing Adversarial Examples, ICLR, 2015.

[5] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, A. Vladu, Towards Deep Learning Models Resistant to Adversarial Attacks, ICLR, 2018.

[6] S. Fujimoto, D. Meger, D. Precup, BCQ: Approximate Bayesian Deep Q-Learning with Behavior Policy Constraints, NeurIPS, 2019.

[7] A. Kumar, A. Zhou, G. Tucker, S. Levine, Conservative Q-Learning for Offline Reinforcement Learning, NeuriPS, 2020.

[8] S. Fujimoto, P. Gupta, A Minimalist Approach to Offline Reinforcement Learning, AAAI, pp. 3571‐ 3579, 2021.

[9] I. Kostrikov, A. N. Nachum, J. Tompson, Offline Reinforcement Learning with Implicit Q-Learning, Proc. ICML, pp. 5774‐ 5783, 2022.

[10] Z. Qian, K. Huang, Q.-F. Wang, X.-Y. Zhang, A survey of robust adversarial training in pattern recognition: Fundamental, theory, and methodologies, Pattern Recognition, Vol. 131, 2022.

[11] L. Pinto, J. Davidson, R. Sukthankar, A. Gupta, Robust Adversarial Reinforcement Learning, ICML, pp.2817‐ 2826, 2017.

[12] J. Tobin, R. Fong, A. Ray et al., Domain Randomization for Transferring Deep Neural Networks from Simulation to the Real World, IEEE/RSJ IROS, pp. 23‐ 30, 2017.

[13] A. Rajeswaran, S. Ghotra, B. Ravindran, S. Levine, EPOpt: Learning Robust Neural Network Policies Using Model Ensembles, ICLR, 2017.

[14] S. Levine, A. Kumar, G. Tucker, J. Fu, Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems, arXiv:2005.01643, 2020.

[15] M. Laskin, A. Srinivas, P. Abbeel, Reinforcement Learning with Augmented Data, Proc. NeurIPS, 2020.
[16] A. Srinivas, M. Laskin, P. Abbeel, CURL: Contrastive Unsupervised Representations for Reinforcement Learning, ICML, pp.9889‐ 9901, 2020.
[17] R. Storn, K. Price, Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces, Journal of Global Optimization, vol.11, pp. 341–359, 1997.
[18] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, O. Klimov, Proximal Policy Optimization Algorithms, arXiv:1707.06347, 2017.
[19] E. Todorov, T. Erez, Y. Tassa, MuJoCo: A physics engine for model-based control, IROS, pp.5026–5033, 2012.
[20] J. Fu, A. Kumar, O. Nachum, G. Tucker, S. Levine, D4RL: Datasets for Deep Data-Driven Reinforcement Learning, arXiv:2004.07219, 2020.

〈発　表　資　料〉

題　　名	掲載誌・学会名等	発表年月
関節トルク信号への摂動に頑健な Offline-to-Online 強化学習	第 39 回人工知能学会全国大会	2025 年 5 月
Decision Transformer による Atari ゲーム AI の画像処理データ拡張評価	第 39 回人工知能学会全国大会	2025 年 5 月
Vulnerability Verification in Robot Control Using Decision Transformer	Proceedings of ISCIIA 2024, Communications in Computer and Information Science, Vol. 2466. pp. 249-261, Springer.	2025 年 4 月
Robustness Verification of Decision Transformer with Varying Noise-Augmented Data Ratios in Atari Games	Proceedings of ISCIIA 2024, Communications in Computer and Information Science, Vol. 2466. pp. 262-273, Springer	2025 年 4 月
Robustness Evaluation of Offline Reinforcement Learning for Robot Control Against Action Perturbations	arXiv:2412.18781	2024 年 12 月
オフライン強化学習におけるデータ拡張による Atari ゲーム AI の性能向上	第 29 回ゲームプログラミングワークショップ予稿集, pp. 87-191	2024 年 11 月
Decision Transformer を用いたロボット制御におけるロバスト性検証	第 38 回人工知能学会全国大会	2024 年 5 月
関節トルク信号への摂動に対するオフライン強化学習手法の頑健性評価	第 38 回人工知能学会全国大会	2024 年 5 月
四脚歩行ロボット体型に対する敵対的攻撃とランダムノイズに頑健な深層強化学習	第 37 回人工知能学会全国大会	2023 年 5 月
四脚アクチュエータに対する敵対的攻撃とランダムノイズに頑健な深層強化学習	第 37 回人工知能学会全国大会	2023 年 5 月