2023 年度 卒業論文

TOKYO DENKI UNIVERSITY

知識選択型転移強化学習を用いた移動ロボットによる動的障害物回避とハイパーパラメータ探査

Dynamic Obstacle Avoidance and Hyperparameter Exploration by a Mobile Robot Using Knowledge-Selective Transitional Reinforcement Learning

指導教員　准教授　河野 仁

東京電機大学 工学部 情報通信学科

学籍番号 20EC070

高矢 空

目次

図目次

表目次

第 1 章

序論

Contents

1.1 背景

1.1.1 ロボットの普及と生活への浸透

ロボット技術の進展は日常生活においても顕著である．その活用範囲は家庭から産業，さらに医療や救助といった特殊な分野まで及んでいる．特に，生活の質を向上させる家庭用ロボットや効率化を促進する産業用ロボットは，私たちの生活に密接に関わっている．これらのロボットは，高度なセンサー技術とアルゴリズムを用いて環境を認識し，タスクを効率的に遂行する能力を有している．昨今，世界中にロボットは日常生活にとって，なくてはならない存在になった．それらのロボットはというものは，国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）の「NEDO ロボット白書 2014」に基づくと，「センサー，知能・制御系，駆動系の 3 つの要素技術を有する，知能化した機械システム」と定義されている．[NEDO 2014]

ロボットは大きく分けると下記の図 1.1 のように「産業用ロボット」と「サービスロボット」の 2 種類に分けられる．例えば，屋内用ロボットは，掃除や料理といった日常の家事を支援することで，私たちの生活を豊かにしている．一方，産業用ロボットは，製造ラインでの精密作業を担い，生産効率の大幅な向上を実現している．また，医療分野では，図 1.2 のような手術支援ロボットが医師の手を補助し，より安全で正確な手術を可能にしている．救助活動においても，災害現場での捜索や救助作業を行うロボットが開発され，人命救助に貢献している．これらのロボットから得られる大量のデータは，技術革新を推進する貴重な情報源である．具体的には，ロボットが収集したデータを分析することで，より効率的なアルゴリズムの開発や，新たな応用技術の創出が可能となる．さらに，これらの技術は，社会のデジタル変革を加速させ，経済や文化の新たな発展を促している．結果として，ロボット技術の普及は，作業効率の向上だけでなく，社会全体のデジタル変革を促進し，新たな価値創造に寄与している．このようにロボット技術は，私たちの生活の質の向上や社会の発展に不可欠な役割を果たしているのである．

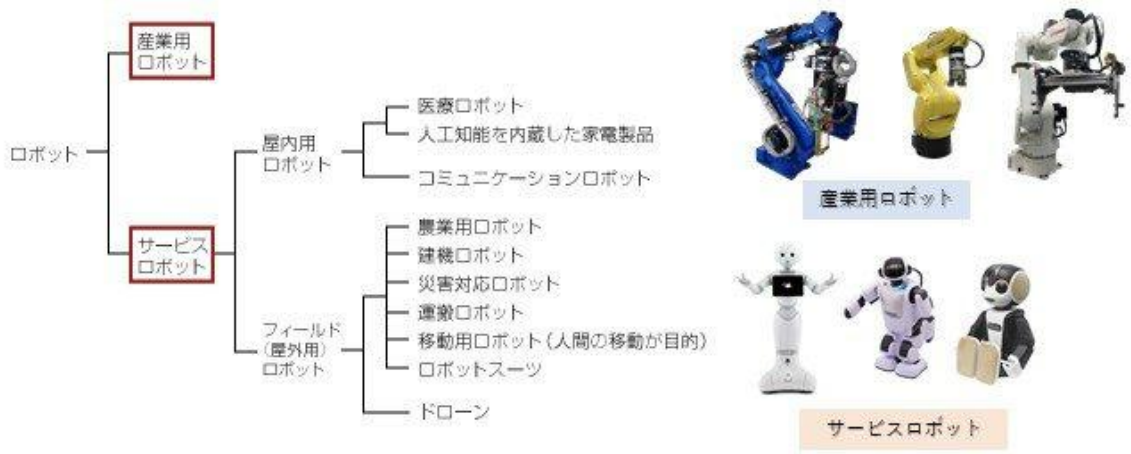


Fig. 1.1: ロボットの分類

Fig. 1.2: 医療ロボットの例

1.1.2 自動運転技術の進展

自動運転技術は，交通の安全性と効率性を根本から変革する大きな可能性を持っている．この技術は，人的ミスに起因する交通事故の減少，交通流の最適化，環境への影響の軽減に大きく貢献することが期待されている．自動運転車は，複数のセンサー技術を活用して周囲の環境を詳細に認識し，先進的なデータ処理技術を用いてこれらの情報を迅速に分析することで，複雑な交通環境においても安全な運転を実現する． 自動運転技術は，自動運転レベルとして政府が定義している．政府の内閣官房情報通信技術総合戦略室は，平成 30 年 4 月に「自動運転係る制度整備大綱」に下記の図 1.3 のように定義している．[IT 総合戦略本部 2018]

自動運転技術の進展は，センサー技術の進化と密接に関連している．ライダー（LIDAR），カメラ，レーダーなどのセンサーは，車両の周囲の状況を 360 度捉えることができ，他の車両や歩行者，さらには突然道に飛び出してくる動物までをも検出する能力を持っている．これらのセンサーから得られる情報は，複雑なアルゴリズムを通じて解析され，自動運転車が安全な運転判断を下せるようになっている． 自動運転車の運転判断の精度を向上させるためには，機械学習やディープラーニングといった先端技術が用いられている．これらの技術により，自動運転車は継続的な学習とデータの蓄積を通じて，その判断能力を日々向上させている．特に，ディープラーニングは，膨大な量のデータから複雑なパターンを学習し，予測することが可能であり，自動運転車がより複雑な環境に対応できるようになるための鍵となっている．さらに，自動運転技術の発展には，車両間通信や車両とインフラの通信技術も重要な役割を果たしている．これらの通信技術により，自動運転車は他の車両や交通インフラと情報を共有し，より安全で効率的な運転が可能になる．例えば，交差点での車両の動きを予測し，衝突を回避するための情報を事前に得ることができる．また，これらの技術の進歩を踏まえ，国土交通省は一般的な道路への自動運転の導入に向けたスケジュールを図 1.4 のように発表しており，2030 年ごろからではあるが，営業運転時期として，営業運転を視野に入れるとしている．

結論として，自動運転技術は，センサー技術，データ処理技術，通信技術の進化に支えられ，交通システムを革新する大きな可能性を秘めている．これらの技術の進展により，利用者の運転負担が軽減されるだけでなく、自動運転車はより安全で効率的な運転を実現し，未来の交通システムに革命をもたらすことが期待される．

レベル	名称	定義概要	安全運転に係る監視、対応主体
運転者が一部又は全ての動的運転タスクを実行			
0	運転自動化なし	運転者が全ての動的運転タスクを実行	運転者
1	運転支援	システムが縦方向又は横方向のいずれかの車両運動制御のサブタスクを限定領域において実行	運転者
2	部分運転自動化	システムが縦方向及び横方向両方の車両運動制御のサブタスクを限定領域において実行	運転者
自動運転システムが（作動時は）全ての動的運転タスクを実行			
3	条件付運転自動化	システムが全ての動的運転タスクを限定領域において実行 作動継続が困難な場合は、システムの介入要求等に適切に応答	システム（作動継続が困難な場合は運転者）
4	高度運転自動化	システムが全ての動的運転タスク及び作動継続が困難な場合への応答を限定領域において実行	システム
5	完全運転自動化	システムが全ての動的運転タスク及び作動継続が困難な場合への応答を無制限に（すなわち、限定領域ではない）実行	システム

Fig. 1.3: 自動運転レベルの定義

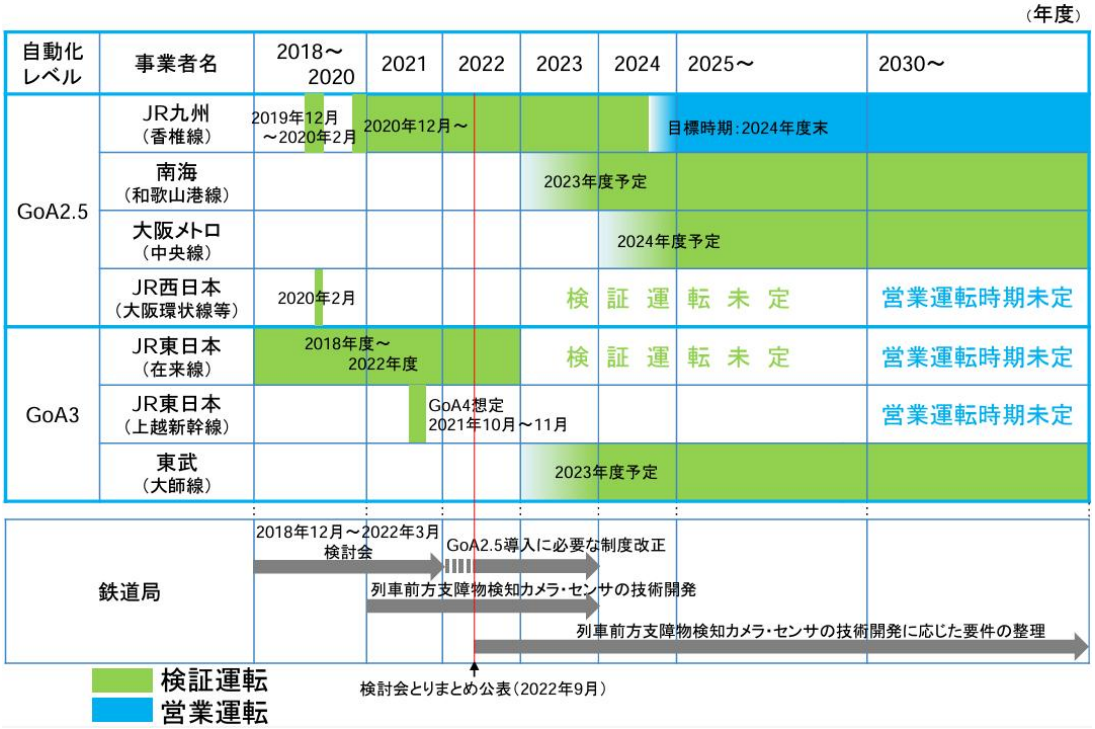


Fig. 1.4: 一般的な道路への自動運転の導入に向けたスケジュール

1.1.3 自動運転技術の事故事例と課題

自動運転技術は大きな進歩を遂げ，多くの利便性をもたらしている．しかし，この技術が完璧ではないことは，過去に発生した事故事例からも明らかである．本稿では，自動運転技術に関連する事故の具体例を挙げ，これらの事故が示す課題について考える．まず，事故事例として，福井県永平寺町での自動運転車事故があげられる．福井県永平寺町で発生した自動運転「レベル 4」の車両事故は，自動運転技術の安全性に対する懸念を示す一例である．図 1.5 のとおり遠隔での監視下にあったにも関わらず，自転車との接触事故を避けることができなかったこの事例は，障害物検知システムの改善が必要であることを示唆している．次に，テスラ車による部分自動運転モード事故である．2016 年 5 月に発生したテスラのモデル S の事故は，部分自動運転モード中に大型トレーラーと衝突し，図 1.6 のような大事故を招き，運転手が死亡した．この事故は，自動運転システムが特定の状況下で障害物を認識しきれない可能性があることを浮き彫りにした．最後に，ウーバーの自動運転車による歩行者死亡事故である．2018 年 3 月，ウーバーの自動運転車が歩行者をはね死亡させる事故がアリゾナ州で発生した．この事故は，図 1.7 のように車体は大破し，燃え上がったという．この事故は自動運転における世界初の歩行者死亡事故とされ，自動運転車が複雑な現実世界の状況に対応する能力の限界を示した．

これらの事故事例からは，自動運転技術にはまだ多くの課題が存在することがわかる．センサーやカメラの限界，学習データの不足，システムの判断過程など，様々な要因が事故発生のリスクを高めているが，いづれも動的障害物回避のアルゴリズムに限界を国土交通省自動車局は示唆している．[国土交通省 2018]

前述した課題に対処するためには，学習データの充実だけでなく，革新的な技術の開発が必要だとされている．革新的な技術を用いることで学習データにない，未知な環境であっても適切に対応をしていく技術が必要だとわかる．

Fig. 1.5: 遠隔監視室を配置した自動運転レベル 4 の区間の図

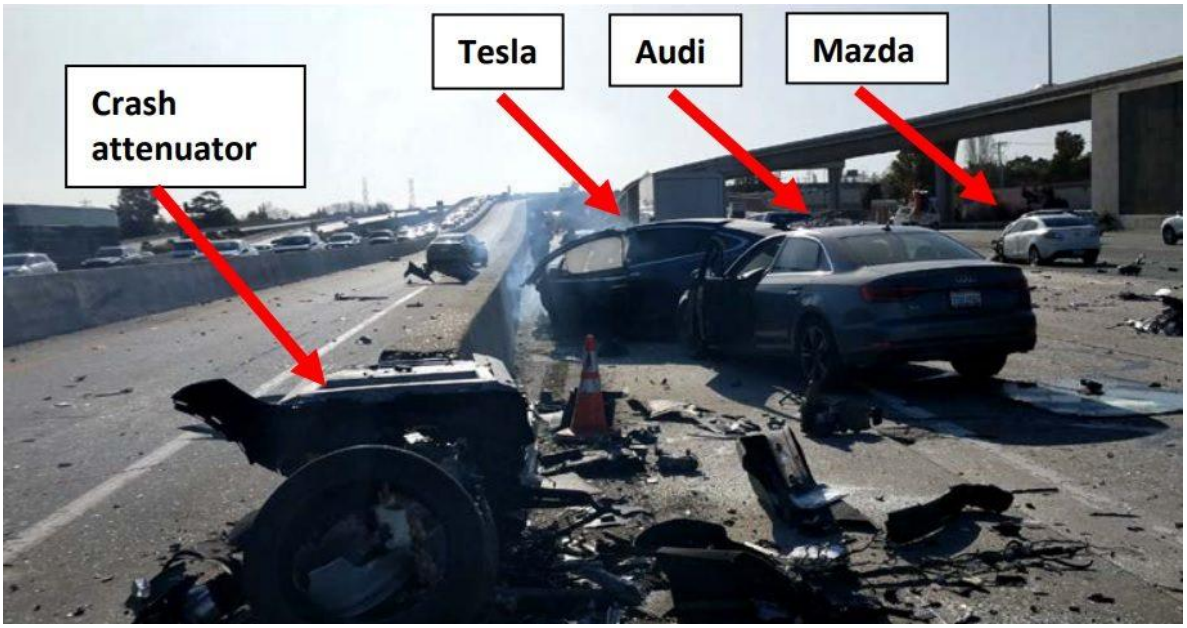


Fig. 1.6: 2018 年 3 月に発生したテスラ車の事故の様子

Fig. 1.7: ウーバーの AI 自動運転車が起こしてしまった死亡事故

1.1.4 動的障害物回避の重要性

動的障害物回避は，自動運転車の性能において最も重要な要素の一つである．交通環境における予測不可能な要素が多い中，自動運転車はこれらの要素に対して迅速かつ正確に対応する必要がある．この能力は，センサー技術の進化，高速なデータ処理能力，そして高度なアルゴリズムの開発によって支えられている．しかし，完全な自動運転の実現に向けては，これらの技術をさらに進化させ，新たな課題に対処するための継続的な研究と開発が不可欠である．実世界の事例を考えると，動的障害物回避の必要性はより明確になる．例えば，突然道路に飛び出してくる歩行者や，予期せぬ場所に停車している車両など，自動運転車はこれらの障害物を検出し，適切な回避行動を取る能力が求められる．このような状況において，先進的なセンサー技術とアルゴリズムは，障害物の正確な位置と動きを迅速に識別し，可能な限り安全な回避経路を計算することが重要である．さらに，既存の研究を引用すると，動的障害物回避技術の発展は，自動運転車の安全性を大幅に向上させることが示されている．例えば，様々なセンサーとアルゴリズムを組み合わせた研究では，複雑な交通環境下での障害物回避性能が向上し，事故のリスクが減少することが報告されている．このような研究成果は，動的障害物回避技術の柔軟に行動できる自動運転システム開発の必要性を強調している．[中川 2005]

また、柔軟に行動できる自動運転システムとは、具体例をあげると、図 1.8 のように、人が歩いているシーンがある。自動運転モビリティが、人を左に検知した。その場合自動運転モビリティは右に回避する。しかし、自動運転モビリティがよけた先にも人が入ってくると、図 1.9 のように衝突してしまうことがわかる。よって、上記のような状況の場合は回避した先でも回避を行うように、柔軟に行動できる自動運転システムの開発を行う必要がある。

動的障害物回避の成功は，自動運転車がより広範囲での安全かつ効率的な運行を実現する鍵となる．この目的を達成するためには，センサー技術，データ処理能力，アルゴリズムの継続的な改善が必要である．したがって，この分野における研究は，将来の自動運転車の安全性と効率性を大幅に向上させる可能性が高い．よって，この課題の緊急性と研究の必要性は，自動運転技術の発展において中心的な役割を果たすと考えられる．

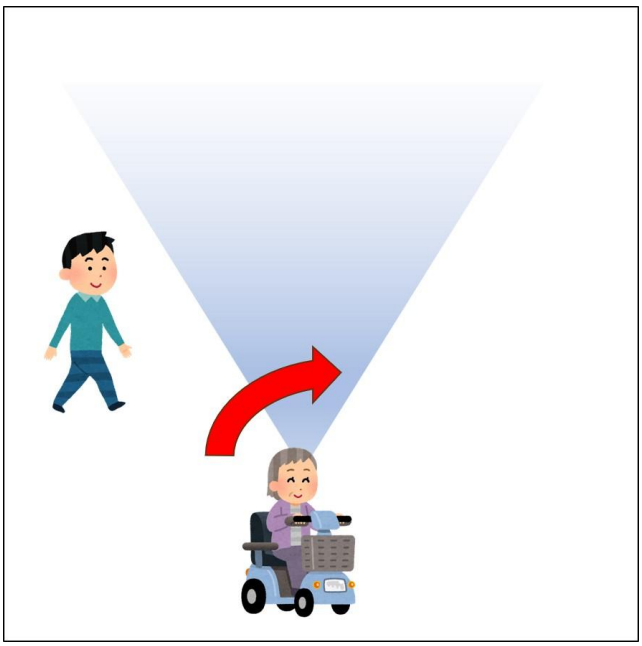


Fig. 1.8: 回避図：回避前

Fig. 1.9: 回避図：回避後

1.1.5　自律型ロボットの導入とその課題

自動運転技術を含む，自律型ロボットの導入は，ロボット技術の発展における次なる大きなステップである．これらのロボットは，人間の直接的な操作や監視を必要とせず，環境内で独立してタスクを実行する能力を持つ．自律型ロボットは，複雑な環境下での決定を行い，未知の状況に適応するために，高度なセンサー技術，人工知能（AI），そして機械学習アルゴリズムを組み合わせて使用している．

自律型ロボットの応用範囲は広く，図 1.10〜図 1.13 のように，災害時の捜索救助活動から，農業，家庭用アシスタント，さらには遠隔地での科学的調査まで多岐にわたる．これらのロボットは，人間に代わって危険な環境で作業を行うことができるため，作業の安全性を大幅に向上させることが可能である．例えば，図 1.14〜図 1.15 のように，原子力発電所での放射性物質の処理や，深海での探査など，人間が容易には行えない作業を自律型ロボットが担うケースが増えている．

しかしながら，自律型ロボットの導入には，いくつかの課題が存在する．技術的な面では，自動運転技術の際に場所の変化への対応や，エージェントの変化に対応できないため，別のエージェントに移行する際にシステム構築を一から行うため，費用が高くなる．また，道の課題に対しては，機械学習の外挿という概念に基づき，予測結果の対応を保証していない．また、日本総研は、2017 年に自動運転が直面する一つの課題に動的な障害物の回避があるとしている。[日本総研 2017]

これらの課題に対応するためには，事前に取得した知識に基づいて AI が自分自身で判断を行い，決断を下す必要がある．

Fig. 1.10: 東北大学開発　レスキューロボット「Quince」

Fig. 1.11: 日本総研開発 自律多機能型農業ロボット「DONKEY」

Fig. 1.12: シャープ開発　家庭用コミュニケーションロボット「RoBoHoN」

Fig. 1.13: リバプール大学開発　自律実験移動型ロボット「Nature」

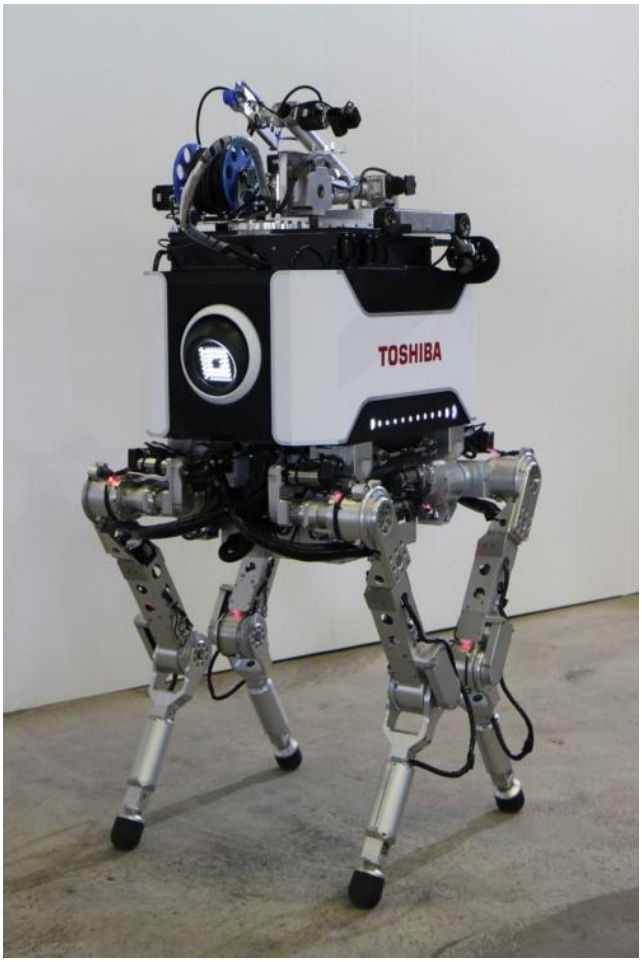


Fig. 1.14: 東芝開発　福島第一原子力発電所廃炉調査ロボット「4 足歩行ロボット」

Fig. 1.15: NASA 開発　深海探査ロボット「Robonaut 2」

1.1.6　課題解決後の自立型ロボットの有用性

前章の課題を解決したロボットは，環境を正確に認識し，適切な判断を下すための次世代的な AI になると考えられる．これらの進展は，自律型ロボットが環境内でのタスクをより効率的に，そして安全に実行する能力を大幅に向上させる．具体的には，先進的なセンサー技術と人工知能（AI）を組み合わせることで，ロボットは未知の環境でも自己位置を把握し，障害物を回避しながら目的地へと移動することが可能となる．さらに，機械学習アルゴリズムの進化により，自律型ロボットは過去の経験から学習し，新たな状況に対しても適切な対応を行うことができるようになる．このような自律型ロボットの能力向上は，人間が行うには危険または困難なタスクを効果的に補助または代行することを可能にする．例えば，災害現場での捜索救助活動では，ロボットはがれきの下で生存者を探索することや，危険な物質が漏れている環境での情報収集を行うことができる．このような状況では，自律型ロボットの高度なセンサーと AI による判断能力が，人命救助の速度と安全性を大幅に向上させる．さらに，農業分野では，自律型ロボットは作物の健康状態を監視し，必要に応じて水や肥料を適切な量で供給することができる．このような精密な管理は，作物の生産性を高めるとともに，資源の使用を最適化し，環境への影響を最小限に抑える．自律型ロボットの導入によって生じるこれらの利点は，社会における生活の質の向上や，産業の効率化に大きく寄与する．しかし，これらの技術革新を社会に広く受け入れるためには，ロボットの行動を予測可能にし，人間との協働を円滑にするためのガイドラインや倫理的な枠組みの確立も必要になってくるだろう．外部の意見を参照すると，国際ロボット連盟（IFR）や IEEE ロボット＆オートメーション協会などの組織は，自律型ロボットの倫理的な使用に関する指針を提案している．これらの指針は，技術開発者や利用者に対して，安全性，透明性，責任のある使用を促すものであり，ロボット技術の社会への統合を促進する上で重要な役割を果たしている．　結論として，課題を解決した自立型ロボットは，多様な分野での応用が期待される．これらのロボットが提供する高度な機能と柔軟性は，人間の生活を豊かにし，産業の発展を加速させるだろう．ただし，その普及には，技術的な挑戦だけでなく，社会的・倫理的な課題への対応も求められる．

1.2 本研究の目的

知識選択型転移強化学習を用いた移動ロボットによる動的障害物回避の実現を目指す．これまでの研究では，静的障害物の回避は実現されているが，動的な障害物の予測と回避は未解決の課題である．本研究では，知識選択型転移強化学習の SAP-net を使用することで動的障害物の回避を行う．さらに，ハイパーパラメータを調整することで，動的障害物の回避をスムーズに実現し，新しい環境への適応度を高めることを目指す．

知識選択型転移強化学習を用いた移動ロボットによる
動的障害物回避とハイパーパラメータ探査

1.3 本論文における研究 3 要素

1.3.1 研究の学術性

本研究は，知識選択型転移強化学習を用いて動的障害物の回避を実現することを提案している．これまでの理論では，静的障害物の回避は可能でしたが，動的障害物に対応する手法は未解決でした．本研究は，この課題に対して新たな視点を提供し，知識選択型転移強化学習によって動的障害物回避の可能性を探ることに学術的な価値がある．

1.3.2 研究の新規性

本研究の新規性は，知識選択型転移強化学習を用いて動的障害物の回避を実現することにある．これまでにないアプローチである知識選択型の転移強化学習を用いることで，動的障害物の予測と回避という従来解決が困難であった課題に対して新たな解決策を提供する．また，ハイパーパラメータの調整を行い，一つの指標を示す点も新規性があると言える．

1.3.3 研究の有用性

本研究の有用性は，実世界の複雑な動的環境において，移動ロボットが効率的に障害物を回避し，安全かつ効率的な運行を実現することにある．自動運転車や災害救助ロボットなど，様々な実用的な応用が期待される．また，学習速度の向上や新しい環境への適応度の向上により，広範な応用シナリオにおいて迅速な対応が可能となる．

1.4　本論文の構成

各章の概要と，論文全体の流れを説明する．

本論文の構成について Fig. 1.16 に示す．本論文は全 5 章から構成されている．第 1 章では，本研究の背景と目的について述べた．

第 2 章では，既存の研究などを調査した結果を述べる．

第 3 章では，第 1 章での内容を踏まえて本研究の提案手法やアプローチについて述べる．

第 4 章では，提案手法の有効性を検証するために行ったシミュレータ実験の内容とその結果について述べる．

第 5 章では，本論文の結論と今後の展望について述べる．

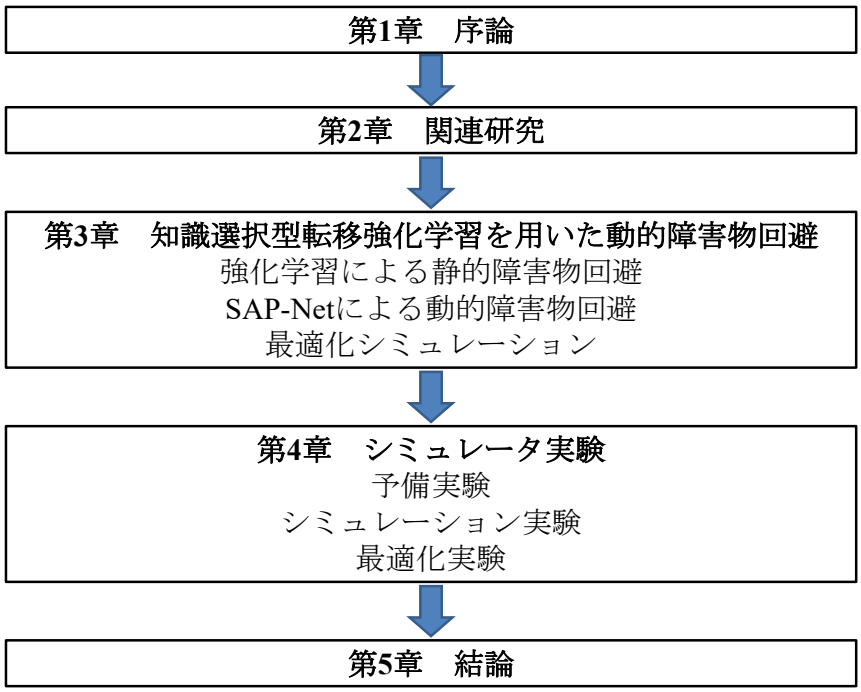


Fig. 1.16: 本論文の構成

第 2 章

関連研究

Contents

2.1 はじめに

本章では，本研究の背景となる関連研究について概説する．動的障害物を回避する移動ロボットの研究は，強化学習，転移学習，およびその他の機械学習技術を用いた多様なアプローチにより進展している．2.2.1 節では，Fernandez の研究する確率分布を転移先タスクでの方策学習に利用する転移学習に関する既存研究について延べる．[Fernandez 2006] 2.2.2 節では，高野らの研究する事前に設定した禁止ルールに基づく転移学習に関する既存研究について延べる．[高野 2011] 2.2.3 節では，河野らの研究する活性化拡散モデルを参考にした知識選択システムによる転移学習に関する既存研究について延べる．[河野 2022]

2.3 節では，2.2.1 節〜2.2.3 節で紹介した既存研究のアプローチとどのように異なるかについても解説する．さらに，動的障害物回避のための既存手法と，それらの手法が直面している課題についても詳細に検討する．これらの関連研究の概説を通じて，本研究の位置づけと，提案するアプローチの革新性及び必要性について明確にする．

2.2 関連研究

2.2.1 確率分布を転移先タスクでの方策学習に利用する転移学習に関する既存研究

確率分布を転移先タスクでの方策学習に利用する転移学習に関する既存研究では，複数のタスクから得られた方策を効果的に活用する方法に焦点を当てている．この研究は，過去に学習した方策が新しいタスクの学習にどのように役立つかを確率的に評価し，適切な方策を選択して再利用することで学習プロセスを加速することを目的としている．具体的には，過去の経験を確率分布として捉え，その分布を新しいタスクの学習に適用することで，効率的な探索と学習のバランスを実現する．このアプローチは，特に似たようなタスク間での知識の移転において，学習の速度と性能の向上を目指し，動的な環境下での適応能力の向上に寄与する可能性がある．このような確率分布を用いた方策学習のアプローチは，転移学習の範囲を拡大し，より複雑なタスクへの適用を可能にするための基盤を提供する．[Fernandez 2006]

また，この研究は最終的に障害物回避はグリッドワールドでの評価のみを行っている．

2.2.2 事前に設定した禁止ルールに基づく転移学習に関する既存研究

この研究は，アクター・クリティック法を用いた転移学習に基づいた学習プロセスの加速を目指している．効率的な転移学習方法を提案し，情報をソースタスクから取得して訓練サイクルを削減する方法，ポリシー選択方法，および各アクター・クリティックパラメーターの特性を考慮した転移方法について述べている．選択方法は，選択フェーズと訓練フェーズの冗長な試行錯誤を削減することを目的としている．また，転移方法は，選択したソースポリシーをターゲットポリシーにマージすることで，負の転移を避けることを目指している．この研究は，シンプルな実験を通じて提案方法の有効性を示している．[高野 2011]

また，この研究は最終的にパスプランニングを使用して，静的障害物回避を行っている．

2.2.3 活性化拡散モデルを参考にした知識選択システムによる転移学習に関する既存研究

この研究は，強化学習における複数の方策から最適な知識を選択する新しい方法を提案している．活性化拡散モデルに基づき，学習済みの方策をカテゴリ化し，状況に応じて適切な方策を選択する．この手法は，特に未知または動的な環境下での学習において，エージェン

トが以前の経験から適切な方策を選択し，学習効率を高めることを目指している．複数の方策を組み合わせることで，エージェントはより柔軟に環境に適応し，既存の単一方策に依存する手法と比較して，学習の効率化と効果の向上が期待される．このアプローチは，様々な強化学習タスクにおけるエージェントの性能改善に寄与する可能性がある．[河野 2022]

また，この研究は最終的に SAP-net を使用して，静的障害物回避を行っている．

2.3 おわりに

本章では，強化学習や転移学習を用いた障害物回避の関連研究について述べた．

2.2.1 節では，Fernandez らの，確率分布を転移先タスクでの方策学習に利用する転移学習に関する既存研究について述べた．

2.2.2 節では，高野らの，事前に設定した禁止ルールに基づく転移学習に関する既存研究について述べた．

2.2.3 節では，河野らの，活性化拡散モデルを参考にした知識選択システムによる転移学習に関する既存研究について述べた．

次章では，転移強化学習を用いた，動的障害物回避の提案手法ついて詳細に述べる．

第 3 章

知識選択型転移強化学習を用いた移動ロボットによる動的障害物回避とハイパーパラメータ探査

Contents

3.1 はじめに

本章では，既存研究である知識選択型転移強化学習を用いた移動ロボットの動的障害物回避手法と，内部で用いるパラメータの調整方法について述べる．3.2 節では，本研究の一連のアプローチ方法を述べる．

3.2 本研究のアプローチ

本研究では，静的障害物回避の知識を強化学習によって獲得する．これらの知識を方策とし，複数の方策を用意する．次に，用意された方策から選択するために，方策ネットワークを活用する．本プロジェクトの第一段階は，方策ネットワークを通じて静的障害物回避の知識を活かして動的障害物を回避することである．その後，回避をより理想的にするために，正しい知識の選択を判断する必要があり，そのためにハイパーパラメータの調整を行う．このアプローチにより，本研究では動的障害物回避の最適な方法を追求していく．

3.3 提案手法に用いる概念

3.3.1 人間の想起に関わる心理モデル

活性化拡散モデルは，認知心理学的なアプローチに基づく想起のメカニズムを表現するモデルであり，ある方策が想起されると，それに関連した方策が活性化され，その活性化が促された方策は想起しやすくなる．このモデルでは，方策同士の関連性が活性化の拡散に影響を与え，関連性が高い方策同士は相互に効果的に想起が促進されるとされている [子安 2011]．これは，関連性の強さに応じて変動する概念間の距離が意味的距離として存在し，これにより関連性の高い概念同士が近くに配置され，効率的な活性化拡散が可能になることを示している．

3.3.2 方策を選択する方策ネットワーク

活性化は関連性によって構築されたネットワークを通じて行われ，各概念間には意味的な関連性の表現が存在する．この概念の活性化によって，関連性がネットワーク上で拡散される仕組みが理解されている．さらに，活性化拡散モデルを学習した複数の方策の関係を記述する手法を用いることで，学習した方策同士においても関連性の表現が可能になり，新たな知識や連想が形成されることが示唆されている．[Collins 1975]．このように，活性化拡散モデルは認知心理学的なアプローチを通じて，方策や概念の想起における関連性とその拡散メカニズムを説明する有益なモデルである．よって，これらの，活性化拡散モデルを学習した複数の方策の関係の記述手法に用いることで，学習した方策同士に関連性の表現が可能になると推測できる [高桑 2017]．図 3.1 に活性化拡散モデルの例を示す．

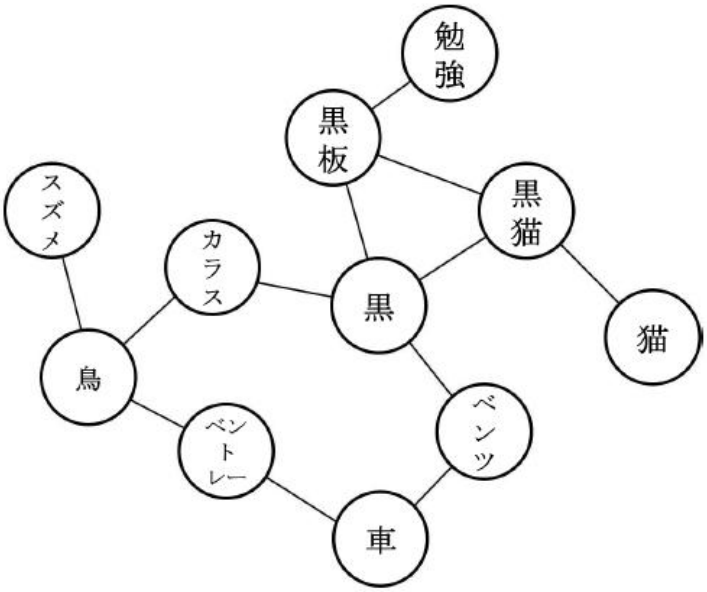


Fig. 3.1: 活性化拡散モデルの例

3.3.3 SAP-net の構成概念

SAP-net とは，3.3.2 で述べた人間の想起に関わる心理モデルをシステム上で構成するために，3.3.2 の，方策を選択するネットワークを用いて，人間の想起を模したシステムを表現する．これにより，表現されるネットワークを SAP-net と呼ぶ.

3.3.4 SAP-net の学術的背景

ロボットシステムに実装される学習アルゴリズムとして，Q 学習やニューラルネットワークなどの様々な手法が提案されている．通常，学習した知識は 1 つだけ生成される．近年ではタスクや環境ごとに知識を個別に保存してネットワーク構造として知識を保存・再利用し，環境適応性を高める研究が行われている．しかし，保存された知識間のネットワークを記述する効果的な手法は確立されていない．一方，人間では概念や意味がネットワーク構造として記憶されているという仮説が様々な心理学実験により確認され，プライミング効果の発現がそれを支持している．プライミング効果とは，先行的に得られる刺激が，後続の処理に無意識的に影響を及ぼす認知心理学の知見である．このプライミング効果が発現するように，ロボットシステム内に保存される獲得知識間のネットワーク構造を記述すれば，知識の検索と選択の正確性と効率を促進できると考える．そこで本研究では，強化学習における複数の知識をネットワーク構造で記述・記憶し，効率な知識選択のために知識ごとに活性レベルを採用する．また，活性拡散モデルに基づいた知識間ネットワークの記述手法と，プライミング効果の発現による知識選択の正確性の向上を目的とする．

また，学習メカニズム自体が研究対象ではなく，知識の保存方法や活性化・想起方法 など，「学習した後」の知識選択手法の確立が目的である． 具体的には，ロボットシステムが様々な環境やタスクを学習し，環境やタスクごとに知識を保 存する．それらの知識に活性値を付加し，センサから先行して得られる情報から知識を活性化さ せることで有効な知識を想起（選択）させるメカニズムの開発を行う．

3.4 提案手法に用いる技術

3.4.1 強化学習の概要とその応用

強化学習の基礎概念と，自動運転車を含む様々な分野での応用例について記載する．

強化学習は，ロボットが試行錯誤を繰り返し，最適な行動を学習していく枠組みである[Sutton 1998][木村 1999]

強化学習は，エージェントが環境内で試行錯誤を繰り返し，最適な行動を学習していく枠組みである．この学習プロセスにおいて，エージェントはある状態において特定の行動を取ることで，環境から報酬というスカラ量を受け取る．エージェントの目的は，受け取る報酬の合計を最大化することにより，目標に適した行動パターンを獲得することである．強化学習のアプローチによって，エージェントは環境に適応し，目標に応じた報酬を最大化する行動を自律的に学習することが可能となる．

強化学習において，多くの研究で採用されている手法の一つが Q 学習である．Q 学習は，ある状態における各行動の期待報酬を表す Q 値を，経験を通じて更新していく手法である．この手法により，エージェントは長期的な報酬を最大化する行動を選択できるようになる．

強化学習は，自動運転車をはじめとする多様な分野で応用されている．自動運転車では，強化学習を用いて，複雑な交通状況の中で安全かつ効率的に目的地に到達するための運転戦略を学習させることができる．また，ロボティクス分野では，強化学習がロボットに自律的な行動をとらせ，未知の環境においても効果的にタスクを遂行させるための手法として利用されている．さらに，強化学習はゲームのプレイや金融市場の取引戦略の最適化など，意思決定が重要な役割を果たす領域でも応用されている．これらの分野では，強化学習が複雑な問題解決において人間の意思決定を補助し，効率化を促進する道具として期待されている．以上のように，強化学習は，エージェントが最適な行動を自律的に学習する枠組みを提供し，自動運転車からロボティクス，ゲーム，金融市場など，幅広い分野においてその有効性が実証されている．

よって、強化学習は、エージェントが環境との相互作用を通じて学習し、目標に向けて最適な行動を決定する手法であるとわかる。この学習プロセスでは、図のように、エージェントは環境からのフィードバックとして報酬を受け取り、行動の良し悪しを評価する。強化学習の目標は、エージェントが未知の状態でも適切な行動を選択できるように学習することである。

代表的な強化学習手法の一つである Q 学習（Q-learning）では、以下の更新式が使用される

$$Q(s,a) \leftarrow (1-\alpha)\cdot Q(s,a) + \alpha \cdot \left(r + \gamma \cdot \max_{a'} Q(s',a')\right) \tag{3.1}$$

ここで、$Q(s, a)$ は状態 s で行動 a を取ったときの行動価値（Q 値）を示し、α は学習率、r は即時報酬、γ は割引率、s' は次の状態を表す。この更新式は、Q 値を現在の値と新しい情報を組み合わせて更新するものである。

強化学習は、様々な応用領域で成功を収めており、実世界の問題に対しても適用が可能である。報酬の最大化を目指すことで、エージェントは複雑な状況においても最適な戦略を学習し、問題を解決する能力を向上させる。

また，ロボットが確率的にある行動をとったときに，目的に合った行動をとると，報酬というスカラ量を得る．学習を進めることで，ロボットは報酬を最大化する行動をとるようになる．つまり強化学習を用いることにより，ロボットの目標に応じた報酬を与えることで環境に適応した動作を学習することができる．本研究では強化学習の中でも多くの研究で用いられている Q 学習を用いる．

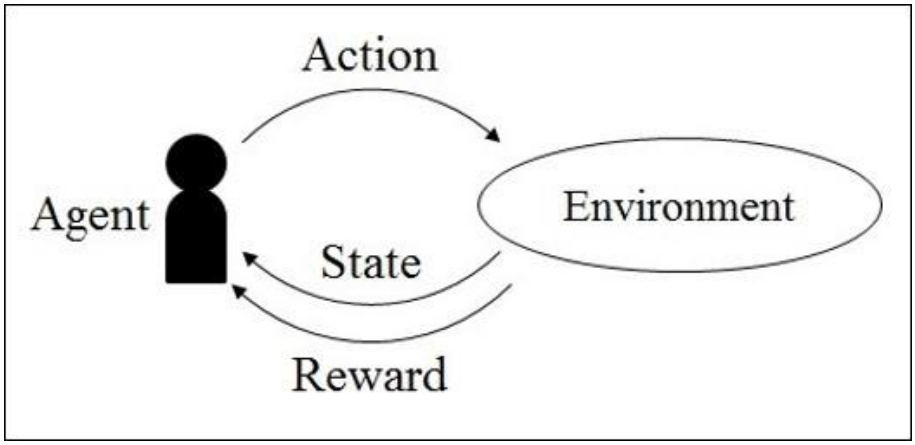


Fig. 3.2: 強化学習の概念図

3.4.2 転移学習の概要とその応用

転移学習（Transfer Learning）とは，あるタスク（ソースタスク）で学習した知識を，別のタスク（ターゲットタスク）へ適用し，学習の効率化を図る機械学習手法である．この手法は，異なるが関連性を有するタスク間で知識を移転することにより，ターゲットタスクの学習に要するデータ量や時間を削減し，モデルの性能を向上させることが可能である．

ソースタスクは転移学習における「知識の供給源」となるタスクである．例えば，大量の画像データを用いて猫と犬を識別するモデルを訓練した場合，この識別タスクはソースタスクとなる．ソースタスクで訓練されたモデルは，画像データの特徴を理解するための豊富な知識を獲得している．対照的に，ターゲットタスクはその獲得した知識を適用したい新しいタスクである．このタスクはソースタスクと何らかの関連性は有するが，完全に同一ではない．例として，ソースタスクで学習した猫と犬の識別モデルを用いて，ライオンとトラを識別するタスクを行う場合，この新しい識別タスクはターゲットタスクとなる．転移学習のプロセスは基本的に以下のステップで構成される．まず，ソースタスクでモデルを訓練する．このステップでは，大量のデータと計算リソースを使用して，モデルがタスクに関連する重要な特徴やパターンを学習できるようにする．次に，訓練されたモデルの一部または全体をターゲットタスクへ転移する．この際，モデルのパラメータの一部を固定（フリーズ）し，残りのパラメータをターゲットタスクのデータで微調整する．この微調整により，モデルはターゲットタスクに特有の特徴を学習しつつ，ソースタスクから転移された知識を保持する．

よって、転移学習は、あるタスクで学習された知識を他の関連タスクに転用する手法であり、機械学習の重要なアプローチの一つであるとわかる。転移学習の目的は、ソースタスク（通常は事前に学習されたモデル）で獲得された知識を、ターゲットタスク（新しいタスク）に効果的に転送することである。このプロセスは、ターゲットタスクにおいて少ないデータで高い性能を達成する上で有益である。転移学習の損失関数は一般的に以下のように表される。

$$L(\theta_s, \theta_t) = \alpha L_s(\theta_s) + (1 - \alpha) L_t(\theta_t) \tag{3.2}$$

ここで、$L(\theta_s, \theta_t)$ は転移学習損失を示し、θ_s と θ_t はそれぞれソースタスクおよびターゲットタスクのパラメータである。また、α はソースタスクとターゲットタスクの寄与を制御する重みであり、通常は 0 から 1 の範囲の値を取る。$L_s(\theta_s)$ および $L_t(\theta_t)$ はそれぞれソースタスクとターゲットタスクにおける損失関数を表す。

この損失関数は、ソースタスクおよびターゲットタスクの損失を組み合わせ、α を通じて

それぞれの寄与を調整することで、効果的な転移学習が可能となる。

転移学習は特にデータが限られている場合や，学習に多大な時間とリソースが必要な場合に有効であり，異なるドメインやタスク間での知識の再利用を可能にするため，様々な分野での応用が期待されている．例えば，医療画像分析，自然言語処理，音声認識など，多岐にわたる分野で転移学習は重要な役割を果たしている．

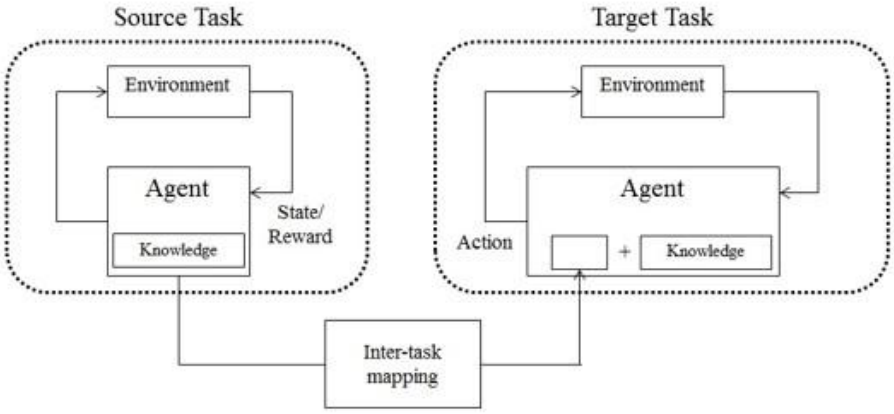


Fig. 3.3: 転移学習の概念図

3.4.3 転移強化学習の概要

転移強化学習（Transfer Reinforcement Learning）は，強化学習と転移学習を組み合わせた手法である．この手法は，あるタスク（ソースタスク）で獲得した知識を別のタスク（ターゲットタスク）に転移させ，学習の効率化と性能向上を図る．

転移強化学習の主な目的は，エージェントが新たなタスクを効率的に学習するために，既存のタスクで取得した知識を活用することである．これにより，エージェントは新たな環境での学習時間を削減し，早期に高性能なポリシーを獲得することが可能となる．

具体的には，転移強化学習は以下のステップで行われる．まず，エージェントはソースタスクで行動を選択し，その結果に基づいて報酬を得て学習する．この過程で，エージェントは状態と行動の間の関係，すなわち Q 値を学習する．次に，学習した Q 値とポリシーをターゲットタスクに転移させる．この際，ソースタスクとターゲットタスクの間で状態空間や行動空間が異なる場合，適切なマッピング関数を用いて Q 値やポリシーを変換する必要がある．その後，エージェントは転移された Q 値やポリシーを基にターゲットタスクでの学習を開始する．この過程で，エージェントは新たな環境に適応し，その環境特有の行動を学習する．

転移強化学習は，新たなタスクへの迅速な適応を可能にし，学習の効率化を図ることができる．しかし，転移の成功はソースタスクとターゲットタスクの間の類似性や関連性に大きく依存する．タスク間の類似性が低い場合，転移による効果は限定的になる可能性がある．そのため，転移強化学習を適用する際には，ソースタスクとターゲットタスクの選択や，適切なマッピング関数の設計などが重要となる．

なお，転移強化学習は強化学習の分野で広く研究されており，ロボティクス，自動運転，ゲームプレイなど，様々な応用例が存在する．これらの分野では，転移強化学習が新たな環境への迅速な適応と，学習の高速化を実現し，効率的な問題解決を支えている．

3.5　提案システムの構成

3.5.1　静的障害物回避の学習

本研究での、第一段階としては強化学習により、知識を複数会得していく。まず、lIDER を用いて図のように、仮想環境に静的障害物を配置する。

今回は図 3.4 のように静的障害物をそれぞれ配置し、強化学習により、それぞれの障害物の回避を強化学習していく。これらの障害物は各回に 1 回ずつ配置され、1 から順に 5 まで配置される。そしてそれぞれの回で強化学習を用いてそれぞれの静的障害物の回避知識を獲得します。また、図 3.4 のロボットモデルが配置されている場所をスタート地点とし、ゴールエリアに入るか、行動回数が打ち切られるまで行動を行う。このスタートからゴールまでのシミュレーションを 4000 回行い、1〜5 の各静的障害物回避を行う。そして、得た強化学習の知識を次章 3.5.2 の動的障害物回避に適用する。

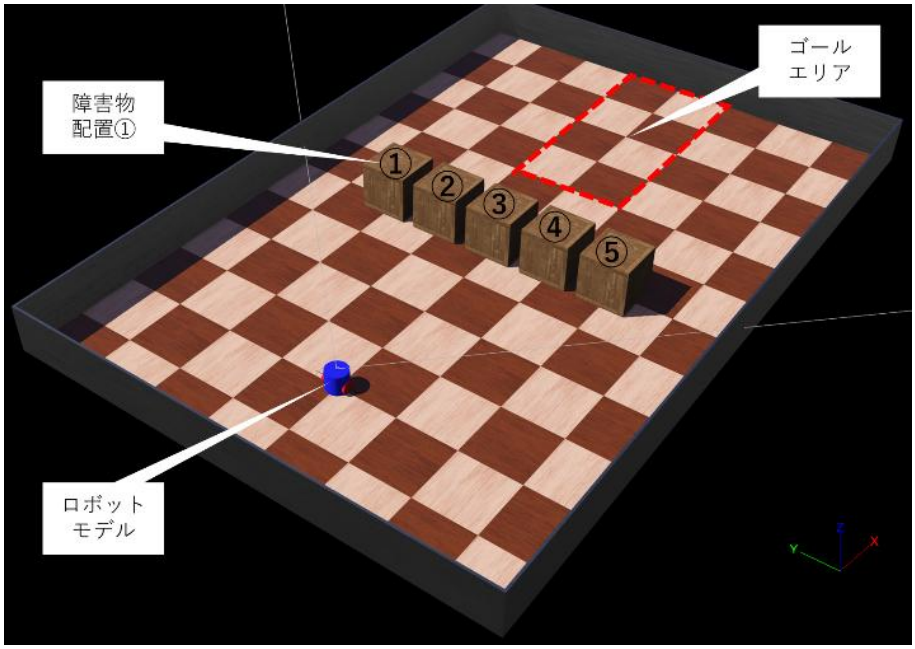


Fig. 3.4: 静的障害物配置例

3.5.2　動的障害物回避への適用

本章では、3.5.1 章で得た静的障害物の回避知識を SAP-net で活用することで動的障害物回避を実現します。動的障害物の配置として、移動前は図 3.5 の状態から、移動後の図 3.6 の状態に移動する。

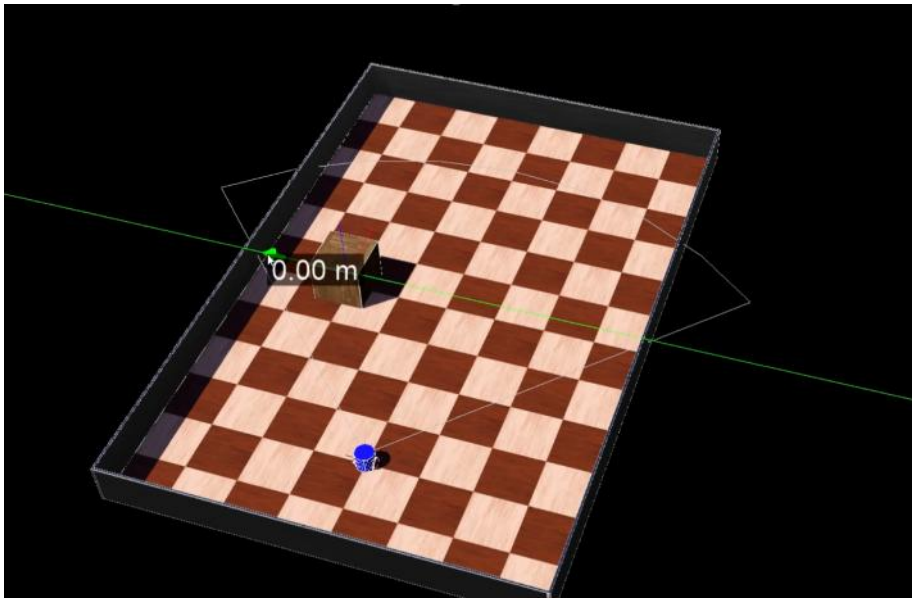


Fig. 3.5: 移動前動的障害物

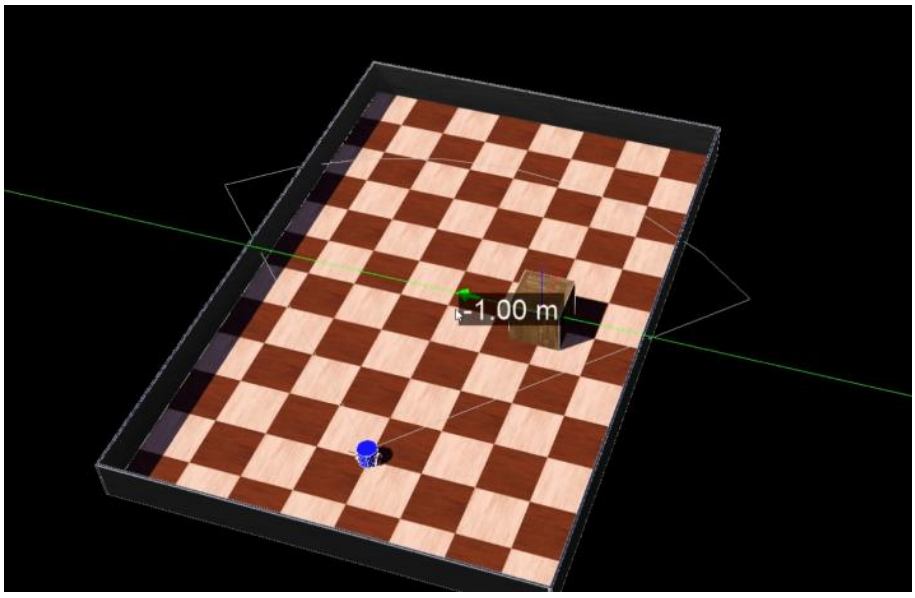


Fig. 3.6: 移動後動的障害物

3.5.3 知識選択型転移強化学習の原理

強化学習で得た知識を方策とし、人間の記憶の想起にまつわる活性化拡散モデルを用いることで、方策を選択する。

SAP-net での方策選択は下記の数式の通りで行われている。

また、SAP-net の詳しい構成や概要などは下記の 3.7 で説明を行う。

3.6 SAP-net の構成

3.6.1 SAP-net の概要

SAP-net は 3.7 のように、環境入力情報とそれぞれの方策に与えられている活性値をもとに，環境に対して効果的な方策の選択を行いる．活性値は環境入力情報により活性化が促され，さらにその活性値が拡散され，閾値により想起が判断される．

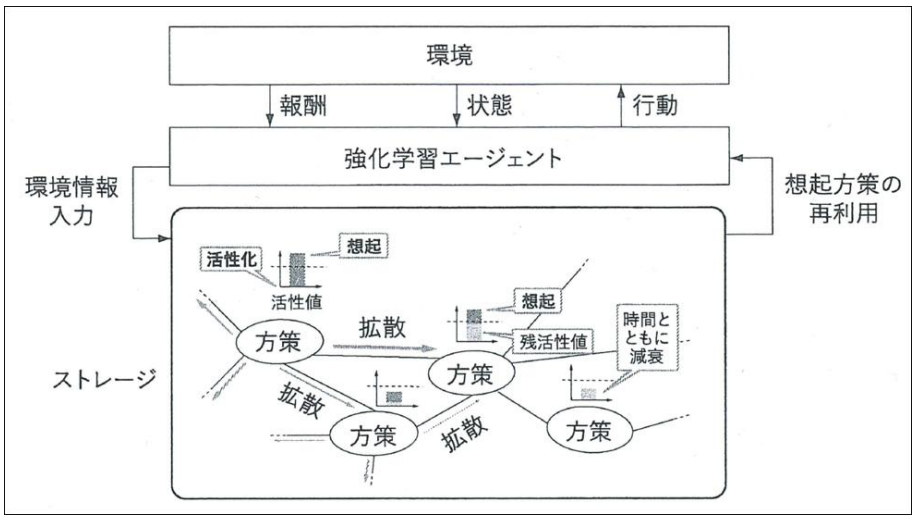


Fig. 3.7: SAP-net の概念図

3.6.2 SAP-net の基本構成

また，SAP-net により，方策がパスで接続され，ネットワーク構造で保存し，人間の脳と同じように活性化拡散を行う．また，下記の手順では，SAP-net に保存された方策から，活性化拡散モデルを用いた転移学習の提案手法の流れを述べる．

1. エージェントは，センサを介して環境情報を観測
2. 観測された環境情報から特徴を抽出し，特徴をラベル化
3. 環境情報の特徴と過去に学習した方策のラベルとが一致する方策を活性化
4. 性化した方策は接続されたパスに対して拡散値を出力
5. 拡散値を受け取った方策はその拡散値から活性値を計算
6. 活性値を持った方策は使用されたパスを除いたパスに対して拡散値を計算
7. すべてのパスを使用するまでステップ 7 を繰り返す
8. 候補方策を活性値をしきい値をもとに収集

9. 確率関数に基づく想起方策選択
10. 方策の転移（この部分は転移学習である.）
11. 学習による行動選択（この部分は強化学習である.）
12. 方策再利用の有効性評価
13. 方策ネットワーク構造の重みの調整
14. ステップ１に戻る

以上の流れにおいて，SAP?net においては，概念を方策に置き換え，環境入力情報とそれぞれの方策に与えられている活性値をもとに，環境に対して効果的な方策の選択が行われている．活性値は環境入力情報により活性化が促され，さらにその活性値が拡散され，閾値にて想起が判断される [坂本 2019]

3.6.3 SAP-net の内部関数

下記に，実際の SAP-net の内部関数をまとめて記す.

- array4DataFrame(array): 配列を DataFrame に変換し，行番号とヘッダーを追加して返す.
- DataFrame4array(df): DataFrame を配列に変換して返す.
- example_data(): サンプルデータを返す.
- example_dataframe(): サンプルデータを含む DataFrame を返す.
- next_Allpair(df, stimulus): 指定された刺激に対して，拡散する先をソートして返す.
- already_pair_remove(pair_list, already_list): 既に存在するペアを除外したリストを返す.
- path_count(df, stimulus): 指定された刺激に対する経路の数をカウントして返す.
- path_weight(df, stimulus, receive): 指定された経路の重みを返す.
- stimulus_pairlist(df, stimulus): 刺激に対する拡散可能なペアのリストを生成して返す.
- stimulus_add_value(path_quantity, path_weight, last_list, pairA, pairB): 刺激値の計算と更新を行い，新しい刺激値と更新後のリストを返す.
- last_dataframe_setting(df, stimulus, first_stimulus_value): 初期刺激値を指定して，新しいデータフレームの最終列を作成する.
- df_update(df, stimulus_value, pairA, pairB): データフレームを更新して返す.
- create_graph(df, GIF_source_path, plotpoint_list): データフレームからグラフを作成し，GIF ソースパスに保存する.

- create_gif(GIF_source_path, GIF_100_path, GIF_1000_path): GIF 用の画像ファイルを生成する．
- attenuation(df, attenuation_percentage): データフレームの値を指定した割合で削減する．
- makeup_folder(): 出力用のフォルダパスと各種ファイルパスを作成して返す．
- create_heatmap(df, Heatmap_path): ヒートマップを作成し，指定されたパスに保存する．
- create_network(df, Network_path): ネットワーク図を作成し，指定されたパスに保存する．
- create_plotpoint(plotpoint_list, Plotpoint_path): プロットポイントを作成する．
- stimulus_calc(df=None, stimulus=1, first_stimulus_value=1.0): サンプルデータを用いて刺激計算を行い，データフレームを返す．

3.6.4 SAP-net のプログラム要件

SAP-net で，実装すべきプログラム要件を下記に示す．SAP-net は今後，別のプログラム内部に組み込む可能性があるため簡易に用いることができるようにするべきである．次に，SAP-net のプログラム要件を以下に示する．SAP-net は将来的に別のプログラム内部に組み込む可能性があるため，簡易に導入できるように設計されている．

プロジェクトの背景と目的の理解

このプロジェクトは SAP-net を用いたプログラムを活用し，適応的なシステムに SAP-net が有用であることを示すことを目的としている．具体的には，少ない学習量でも人間が応用できる知識をシステムに統合し，静的障害物回避の知識を強化学習で取得することで，動的障害物の回避が可能となる．

ステークホルダーの識別

プロジェクトには，今後の動的障害物の回避に関心を寄せるステークホルダーが含まれている．プログラムは簡単に導入可能である必要があり，そのためにはパッケージ化が行われている．

ユーザー要件の収集

ユーザー要件は，Lider にプログラムを導入して動的障害物の回避に活用することが求められる．

機能要件の定義

システムは SAP-net の知識を活用して，重み，距離，角度などを計算し，適切な知識を選択する機能を提供する．以下は主な機能だ．

- `array4DataFrame(array)`: 配列を DataFrame に変換し，行番号とヘッダーを追加して返す．
- 他の機能も含め，DataFrame や配列の変換，データ処理，グラフ作成などが含まれる．

非機能要件の定義

プログラムはリアルタイムで動的障害物の回避判断を行い，高速かつ安定したレスポンスを提供する必要がある．将来的にはこれらの要件が絶対的に必要となるが，今年度は課題としての取り組みが可能だ．

データ要件の確定

システムが取り扱うデータは SQL と CSV でローカルに保存され，適宜出力される．データの種類は Python のデータフレームで扱われ，別プログラムに渡す場合は SQL で処理される．また，人間的な処理が入る場合は CSV として出力し，疑似的な待機が可能となっている．

システムの制約と制約条件

予算は０円であり，約１年間の制約がある．技術要件としては，どのＯＳにも依存せず利用可能なプログラムを開発する必要がある．

ユーザビリティとユーザインターフェースの要件

プログラムは pypi にアップロードされており，pip でダウンロードできるようになっている．一行で呼び出せるように関数化され，簡単に導入可能だ．

テスト要件の定義

テスト要件は手計算での結果確認，さまざまな知識の確認，および P-SAP-net との一致を含む．

プロジェクトスケジュールと進捗管理

プロジェクトは 8 月に北海道で論文発表があり，そのためには動的障害物の回避が必要だ．卒業までには動的障害物の回避とハイパーパラメータの調整が必要だ．進捗管理として，マイルストーンを設定し，スケジュールに従って進捗を確認する．

3.6.5 SAP-net の速度上限

3.6.6 SAP-net の使用方法

3.7 減衰関数について

3.8 数値的な評価指標

3.9 最適化シミュレーション

最適化技術，および性能向上のための戦略について書く．

3.10 提案システムの全体像

パーツの構成全体を見せる

3.11 おわりに

本章では，リカバリモーション獲得の詳細について述べた．

??節では，本研究で用いるシミュレータ内のロボットコントローラと強化学習器の関係について述べた．

??節では，本研究で用いる強化学習のうち Q 学習について述べた．また，Q 値に関して RBF ネットワークによる近似を行うことを述べた．

??節では，強化学習内で用いる報酬関数について述べた．本研究で用いる報酬関数として移動ベクトルの報酬，目標到達の報酬を設定し，学習の促進を図る．また，ロボットの移動中の安定性を評価するために，報酬関数に NE 安定余裕の報酬を設定した．

次章では，動力学シミュレータを用いた実験を行う．

第 4 章

シミュレータ実験

Contents

4.1 はじめに

本章では，提案したアプローチの有用性を示すために行った，動力学シミュレータ実験について述べる．

4.2 節では，本研究で用いるロボットとシミュレータ環境について述べる．

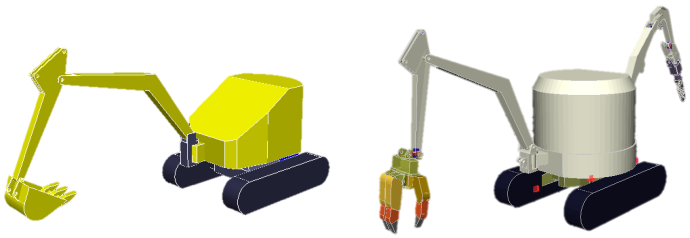
??節では，単腕クローラロボットによる直進と旋回のリカバリモーション獲得の実験について述べる．

??節では，手法が他のアーム搭載ロボットにも適用可能かを検証するために，双腕クローラロボットによるリカバリモーション獲得の実験について述べる．

最後に 4.10 節で本章のまとめを述べる．

4.2 実験環境について

本研究で用いるロボットは Fig. 4.1(a) と Fig. 4.1(b) に示す単腕クローラロボットと双腕クローラロボットである．単腕クローラロボットはアーム 1 本を搭載したクローラ型ロボットである．双腕クローラロボットはアーム 2 本を搭載したクローラ型ロボットである．ロボットから見て左のクローラの故障を想定する．Fig. 4.2 に示すようにロボットの移動は平らな水平面で行われ，周りに障害物はないものとする．単腕クローラロボットに関しては，直進，旋回のリカバリモーションを獲得する．双腕クローラロボットに関しては，直進のリカバリモーションを獲得する．



(a) Single-armed crawler robot　(b) Double-armed crawler robot

Fig. 4.1: Single-armed and double-armed crawler robot

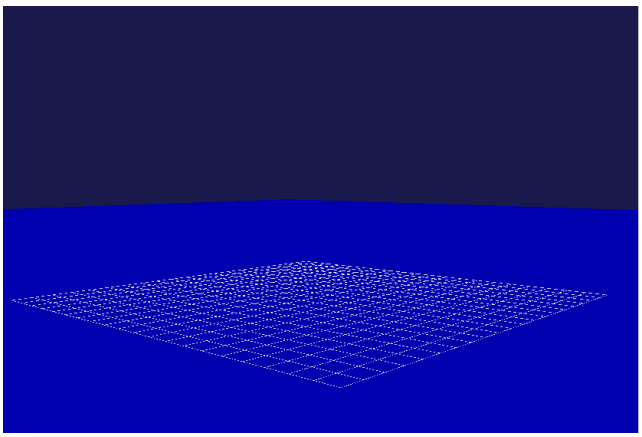


Fig. 4.2: Flat surface in Choreonoid 1.5

4.3 予備実験

実際の実験に用いるであろう技術の環境を準備しテストした

4.3.1 予備実験目的

4.3.2 予備実験条件

実験条件とシミュレーション環境

4.3.3 予備実験結果

4.4 静的障害物回避実験

4.4.1 静的障害物回避実験目的

4.4.2 静的障害物回避実験条件

実験の設計，使用するシミュレーション環境，および評価指標について説明する

4.4.3 静的障害物回避実験評価

実験の目的，手法，および評価基準について詳述する（プログラムの動作する秒数やアルゴリズム的に実装が可能な点．論理的整合性に間違いがない点を示す．

4.4.4 静的障害物回避実験結果

実験結果の詳細と，それらの結果がどのように提案手法の有効性を示すかについて分析し，定量的に実験の結果を示す

4.5 動的障害物回避実験

4.5.1 動的障害物回避実験目的

4.5.2 動的障害物回避実験条件

実験の設計，使用するシミュレーション環境，および評価指標について説明する

4.5.3 動的障害物回避実験評価

実験の目的，手法，および評価基準について詳述する（プログラムの動作する秒数やアルゴリズム的に実装が可能な点．論理的整合性に間違いがない点を示す．

4.5.4 動的障害物回避実験結果

実験結果の詳細と，それらの結果がどのように提案手法の有効性を示すかについて分析し，定量的に実験の結果を示す

4.6 知識選択実験

4.6.1 動的障害物回避実験目的

4.6.2 動的障害物回避実験条件

実験の設計，使用するシミュレーション環境，および評価指標について説明する

4.6.3 動的障害物回避実験評価

実験の目的，手法，および評価基準について詳述する（プログラムの動作する秒数やアルゴリズム的に実装が可能な点．論理的整合性に間違いがない点を示す．

4.6.4 動的障害物回避実験結果

実験結果の詳細と，それらの結果がどのように提案手法の有効性を示すかについて分析し，定量的に実験の結果を示す

4.7 最適化実験

SAP-net の応用とハイパーパラメータチューニングを行う．

4.7.1 最適化実験目的

最適な性能を得るためのハイパーパラメータの調整方法について書いてみたい．（未定）

4.7.2 最適化実験条件

4.7.3 最適化実験結果

提案手法の有効性と限界

4.8 実験結果

提案手法の強みと，現時点での限界について考察し，今後の展望につなげられる文を書く

4.9 手法比較

提案手法と他の既存手法との比較に基づいた評価を行ってみる．

4.10 おわりに

本章では，強化学習を用いたリカバリモーションの獲得のアプローチについての有効性を検証するために行った動力学シミュレータ実験について述べた．

本実験で用いたクローラロボットとシミュレータの環境について述べた．

??節では，本研究のアプローチを左クローラが故障している単腕クローラロボットの直進と旋回のリカバリモーションについて行い，直進，旋回のリカバリモーションの獲得が可能であることを示した．学習が進むことで無駄な動きが減り，目標の移動方向への移動が可能となった．また，得られた 2 つのリカバリモーションは安定性を保った移動方法であることを確認した．

??節では，双腕クローラロボットの直進のリカバリモーションについて行い，直進リカバリモーションの獲得が可能であり，単腕クローラロボット以外にも本手法が適用できることを示した．

第 5 章

結論

Contents

5.1 結論

研究全体の要約と，達成された主要な成果を強調する．（電気学会のものを軸に書いてみる）

本研究ではアーム搭載クローラロボットにおける，故障時の移動方法であるリカバリモーションを獲得する枠組みを提案した．

第 1 章では，本研究の背景となる遠隔操作ロボットの故障時における問題点と，その解決方法として新たな移動方法であるリカバリモーション提案した．また，災害現場において活用する点，遠隔操作ロボットで用いる点，関連する研究を踏まえ，リカバリモーションに関する条件を述べ，本研究の目的を「安定性を考慮したクローラ故障時におけるアーム搭載ロボットのリカバリモーションの獲得」とした．

第 2 章では，第 1 章で述べた課題設定とリカバリモーションのアプローチをもとに必要な条件などを整理し，システムのコンセプトを述べた．また，本研究で必要となる強化学習の基礎知識について述べた．第 3 章では，強化学習の詳細な構成について，第 2 章で述べた必要な条件を踏まえて述べた．報酬関数として移動ベクトル，ゴール報酬に加え，NE 安定余裕に基づいたロボットの安定性の報酬を加えることを述べた．

第 4 章では，提案手法の有効性を確認するために動力学シミュレータによる実験を行った．本研究のアプローチを左クローラが故障している単腕クローラロボットの直進と旋回のリカバリモーションについて行い，直進，旋回のリカバリモーションの獲得が可能であることを示した．双腕クローラロボットの直進のリカバリモーションについても行い，直進リカバリモーションの獲得が可能であり，単腕クローラロボット以外にも本手法が適用できることを示した．また，得られたリカバリモーションは安定性を保った移動方法であることを確認した．実験結果から，本研究で設定した報酬設計を用いることで，強化学習によりクローラ故障時における安定性のあるリカバリモーションの獲得が可能であった．

以上から，本研究における提案手法の有効性が示された．

5.2 今後の展望

今後の研究の方向性と，提案手法のさらなる応用可能性について述べる．
今後の展望としては，以下の 3 点が考えられる．

1. 学習アルゴリズムのパラメータチューニング．
2. 不整地面でのシミュレータ実験．
3. 得られた移動モーションの実機での確認．

今回行った実験では学習の収束に多くの時間がかかった．そこで学習アルゴリズムに関係するパラメータチューニングが必要である．さらにロボットごと，状況毎に異なるパラメータが必要であると考えられるが，このパラメータについても最適化する枠組みが必要であると考えられる．

また，本研究は不整地で活躍する災害対応ロボットを想定している．本手法で用いた NE 安定余裕は不整地においても正しく安定性を評価できる概念である．そのため凹凸のある地面や斜面の環境に関して追加でシミュレータ実験を行う必要がある．またシミュレータ上ではアームの先端が地面で滑らないことも可能性として考えられるので，この条件に関してもシミュレータ実験を行う必要がある．以上の実験を重ねることでロボットの安定性を考慮したことのさらなる有効性が示される．

本研究では実機にて発揮すべきシステムである．そのため，最終的には得られたリカバリモーションが実機において実行可能かの検証が必要となる．本研究で得られたリカバリモーションはアームによってロボットを支える動作であった．そのためアームの可動範囲，ロボットごとのアームの耐久性を考慮する必要がある．また，実行不可能である場合，どのような改良・制約を加えれば実現可能なリカバリモーションが可能となるのか比較検討が必要となる．

謝辞

本論文を締めくくるにあたり，ご指導，ご協力をいただいた全ての方々に，深く感謝いたします．

本研究の指導教員である東京大学大学院 工学系研究科 精密工学専攻准教授 山下 淳 先生には，有意義な研究の機会を与えていただくとともに，熱心なご指導を賜りました．研究論文や発表資料の添削をして下さっただけでなく，研究に関する様々な疑問に対して，いつも非常に納得できる説明や助言をして下さり，ものを考える力を鍛えて下さいました．この経験は非常に有意義で，今後の人生に大きく役立つと確信しています．ここに深く感謝いたします．

東京大学人工物工学研究センター教授 太田 順 先生には，本論文をご精読頂き，有益なご指摘・ご助言を頂きました．ここに深く感謝いたします．

東京大学大学院 工学系研究科 精密工学専攻教授 淺間 一 先生には，ご多忙のなかご指導ご鞭撻を賜り，研究に対する考え方の多くを学ばせていただきました．学術的なレベルの高さだけでなく，論理の適切さ，研究に対する真摯さをご教授賜り，研究を進めていくにあたって非常に勉強になりました．ここに深く感謝いたします．

東京大学大学院 工学系研究科 精密工学専攻特任准教授　田村 雄介 先生には研究論文・発表資料の添削や研究発表におけるご指導を賜りました．誠にありがとうございます．

東京大学大学院 工学系研究科 精密工学専攻特任研究員 河野 仁 博士には直接指導を受け，さまざまな研究の知識やテクニックを教えていただきました．特に災害ロボット，学習という分野に関して大変お世話になりました．また，研究生活においても非常に親身にして下さり，非常に充実したものとなりました．誠にありがとうございます．

東京大学大学院 工学系研究科 精密工学専攻特任助教 安 先生には，異なる研究分野であるからこそ，違った視点でのものの見方を学ぶことができました．誠にありがとうございます．

東京大学大学院 工学系研究科 精密工学専攻特任助教 藤井 浩光 先生には，研究発表のご指導を賜っただけでなく，研究生活における様々な相談に乗っていただいたり，助けていただいたりしました．ここに心より感謝申し上げます．

東京大学大学院 工学系研究科 精密工学専攻特別研究員 池 勇勳 博士には本論文の添削において有益なご指摘・ご助言を頂きました．また，グループミーティングでは的確なアドバイスをいただきました．誠にありがとうございます．

東京大学大学院 工学系研究科 精密工学専攻技術専門員 山川 博司 先生には研究生活を送るうえで重要な身の回りの環境を整えてくださいました．誠にありがとうございます．

石川 雄己氏をはじめとする研究室の先輩方には，研究発表のご指導を賜っただけでなく，研究生活における様々な相談に乗っていただきました．ここに深く感謝いたします．

同じ B グループとして研究を行った淺間・山下研究室の Woo Hanwool 氏，Miyagusuku Renato 氏，陸 小軍氏，田中 佑典氏，金 渡演氏，邵 宇陽氏，Mai Ngoc Trung 氏，江 君氏，Seow Yip Loon 氏とは，日頃より研究に関して議論を交わし，研究を進めていくうえで，重要なヒントを得たり，諦めず継続して行うことを学んだり，有意義な時間を過ごさせていただきました．大変感謝いたします．

同輩である杉本 賢勇君，吉田 和憲君，奥村 有加里さんとは，研究に関して励ましあったり，刺激を受けあったりしただけでなく，研究室での生活全般に関して非常に有意義に過ごすことができました．心から感謝いたします．

秘書の成島 久恵さん，小島 里佳さん，中村 惠さん，石田 万紀さん，後藤田 彩さんには，研究活動を行う上で必要な事務手続きなどの業務を円滑に行っていただいたおかげで，集中して研究をこなすことができました．誠にありがとうございます．

最後に，私の大学での学びを経済的，精神的に支えてくれた家族，そして友人の方々に深く感謝いたします．本当にありがとうございました．

平成 29 年 2 月　伊藤翼

参考文献

＜和文文献＞

[NEDO 2014]
国立研究開発法人新エネルギー・産業技術総合開発機構（NEDO）: “NEDO ロボット白書 2014,” 2014.

[日本総研 2017]
自動運転車の五つの社会実装課題に対する技術的なアプローチの提案: 2017.

[IT 総合戦略本部 2018]
IT 総合戦略本部: “自動運転に係る制度整備大綱,” 2018.

[国土交通省 2018]
国土交通省自動車局: “自動運転における損害賠償責任に関する研究会,” 2018.3.

[中川 2005]
中川　真仁: “動的障害物回避に注目した電動四輪車の知的自動運転システム,” 日本知能情報ファジィ学会誌,2005.

[青柳 2021]
青柳　誠司: “移動ロボットの移動障害物回避に関するファジィルールの学習?ポテンシャル法，強化学習法との比較,” システム制御情報学会論文誌, 2021.

[Fernandez 2006]
Fernando Fernandez: “移確率分布を転移先タスクでの方策学習に利用する転移学習,” 5th International Joint Conference on Autonomous Agents and Multi-agent Systems, 2006.

[高野 2011]
高野 敏明: “アクタークリティカル法で設定された禁止ルールに基づく転移学習,” International Journal of Innovative Computing, Information & Control, 2011.

[河野 2022]
河野 仁: “活性化拡散モデルを参考にした知識選択システムによる転移学習,” 科学研究費助成事業 研究成果報告書, 2022.

[子安 2011]
子安 増生: “認知心理学,” 新曜社出版, 2011.

[Collins 1975]
Collins, Allan M.: “意味処理の拡散活性化理論,” APA PsycArticles Journal Article, 1975.

[高桑 2017]
高桑 優作: “活性化拡散モデルに基づく強化学習エージェントの方策選択手法,” ロボティクス・メカトロニクス講演会,2014.

[坂本 2019]
坂本　裕都: “転移学習を用いた強化学習ロボットにおける方策選択の認知的経済性の検討,” ロボティクス・メカトロニクス講演会,2019.

[坂本 2019]
坂本　裕都: “転移学習を用いた強化学習ロボットにおける方策選択の認知的経済性の検討,” ロボティクス・メカトロニクス講演会,2019.

[木村 1999]
木村 元: “強化学習システムの設計指針,” ロボティクス・メカトロニクス講演会,1999.

[Sutton 1998]
R.S.Sutton: “強化学習,” 森北出版,1998.

？

＜英文文献＞

[Matsuno 2004]
F. Matsuno and S. Tadokoro: “Rescue Robots and Systems in Japan,” *Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics*, pp. 12–20, 2004.

[Murphy 2004]
R. R. Murphy: “Trial by Fire [Rescue Robots],” *IEEE Robotics & Automation Magazine*, vol. 11, no. 3, pp. 50–61, 2004.

[Carlson 2005]
J. Carlson and R. R. Murphy: “How UGVs Physically Fail in the Field,” *IEEE Transactions on Robotics*, vol. 21, no. 3, pp. 423–437, 2005.

[Messuri 1985]
D. A. Messuri and C. A. Klein: “Automatic Body Regulation for Maintaining Stability of a Legged Vehicle during Rough-terrain Locomotion,” *IEEE Journal on Robotics and Automation*, vol. 1, no. 3, pp. 132–141, 1985.

[Platt 1991]
J. Platt: “A Resource-Allocating Network for Function Interpolation,” *Neural Computation*, vol. 3, no. 2, pp. 213–225, 1991.

[Nagatani 2011]
K. Nagatani, S. Kiribayashi, Y. Okada, S. Tadokoro, T. Nishimura, T. Yoshida, E. Koyanagi and Y. Hada: “Redesign of Rescue Mobile Robot Quince,” *Proceeding of 2011 IEEE International Symposium on Safety, Security, and Rescue Robotics*, pp. 13–18, 2011.

[Kawatsuma 2012]
S. Kawatsuma, M. Fukushima and T. Okada: “Emergency Response by Robots to Fukushima-Daiichi Accident: Summary and Lessons Learned,” *Industrial Robot: An International Journal*, vol. 39, no. 5, pp. 428–435, 2012.

[Kober 2013]
J. Kober, B. J. Andrew and Jan Peters: "Reinforcement Learning in Robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[Haykin 2009]
S. Haykin: Neural Networks and Learning Machines, Pearson Upper Saddle River, 2009.

[Mnih 2013]
V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra and M. Riedmiller: "Playing Atari with Deep Reinforcement Learning," *Proceedings of NIPS 2013Deep Learning Workshop*, 2013.

研究業績

査読有り国内会議

1. **伊藤 翼**, 河野 仁, 田村 雄介, 山下 淳, 淺間 一: “アーム搭載移動ロボットの駆動系故障時のための強化学習を用いたリカバリモーション獲得,” 第 22 回ロボティクスシンポジア予稿集, 2017, 発表予定.

査読有り国際会議

1. **Tasuku Ito**, Hitoshi Kono, Yusuke Tamura, Atsushi Yamashita, and Hajime Asama: “Recovery Motion Learning for Arm Mounted Mobile Crawler Robot in Drive System’s Failure,” *The 20th World Congress of the International Federation of Automatic Control*, 2017, 査読中.