

2023 年度 卒業論文



情報エントロピーを用いた知識選択型転移強化学習 のハイパーパラメータ探査と評価方法の開発

Hyperparameter Exploration and Evaluation Methods for
Knowledge-Selective Transitional Reinforcement Learning Using
Information Entropy

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

第 1 章	序論	1
1.1	背景	2
1.1.1	ロボットの普及と生活への浸透	2
1.1.2	自動運転技術の進展	3
1.1.3	自動運転技術の事故事例と課題	5
1.1.4	動的障害物回避の重要性	6
1.1.5	自律型ロボットの導入とその課題	6
1.1.6	課題解決後の自立型ロボットの有用性	7
1.2	本研究の目的	8
1.3	本論文における研究 3 要素	9
1.3.1	研究の学術性	9
1.3.2	研究の新規性	9
1.3.3	研究の有用性	9
1.4	本論文の構成	10
第 2 章	関連研究	11
2.1	はじめに	12
2.2	関連研究	13
2.2.1	強化学習を用いたロボットの行動獲得に関する既存研究	13
2.3	おわりに	14
第 3 章	提案手法名をここに書く	15
3.1	はじめに	16
3.2	本研究のアプローチ	17
3.3	提案手法に用いる技術	17
3.3.1	強化学習の概要とその応用	17

3.3.2	転移学習の概要とその応用	18
3.3.3	転移強化学習の概要	18
3.4	提案システムの構成	19
3.4.1	動的障害物回避への適用	19
3.4.2	知識選択手法の概念	19
3.4.3	知識選択型転移強化学習の原理	19
3.4.4	SAP-net の概念	19
3.5	SAP-net の構成	20
3.5.1	SAP-net の基本構成	20
3.5.2	SAP-net の内部関数	20
3.5.3	SAP-net のプログラム要件	20
3.5.4	SAP-net の速度上限	20
3.5.5	SAP-net の使用方法	20
3.6	数値的な評価指標	20
3.6.1	減衰関数について	20
3.7	最適化シミュレーション	20
3.8	提案システムの全体像	20
3.9	おわりに	21
第 4 章	シミュレータ実験	23
4.1	はじめに	25
4.2	実験環境について	26
4.3	予備実験	27
4.3.1	予備実験目的	27
4.3.2	予備実験条件	27
4.3.3	予備実験結果	27
4.4	静的障害物回避実験	28
4.4.1	静的障害物回避実験目的	28
4.4.2	静的障害物回避実験条件	28
4.4.3	静的障害物回避実験評価	28
4.4.4	静的障害物回避実験結果	28
4.5	動的障害物回避実験	29
4.5.1	動的障害物回避実験目的	29
4.5.2	動的障害物回避実験条件	29

4.5.3	動的障害物回避実験評価	29
4.5.4	動的障害物回避実験結果	29
4.6	知識選択実験	30
4.6.1	動的障害物回避実験目的	30
4.6.2	動的障害物回避実験条件	30
4.6.3	動的障害物回避実験評価	30
4.6.4	動的障害物回避実験結果	30
4.7	最適化実験.....	31
4.7.1	最適化実験目的.....	31
4.7.2	最適化実験条件.....	31
4.7.3	最適化実験結果.....	31
4.8	実験結果	31
4.9	手法比較	31
4.10	おわりに	32
第 5 章	結論	33
5.1	結論	34
5.2	今後の展望.....	35
謝辞		39
参考文献		41
研究業績		46

目次

1.1	ロボットの分類.....	3
1.2	自動運転レベルの定義.....	4
1.3	The structure of this thesis	10
4.1	Single-armed and double-armed crawler robot	26
4.2	Flat surface in Choreonoid 1.5	26

表目次

2.1	Expectation of fire departments to robotic systems [田所 2012]	14
-----	--	----

第 1 章

序論

Contents

1.1	背景	2
1.1.1	ロボットの普及と生活への浸透.....	2
1.1.2	自動運転技術の進展	3
1.1.3	自動運転技術の事故事例と課題.....	5
1.1.4	動的障害物回避の重要性	6
1.1.5	自律型ロボットの導入とその課題	6
1.1.6	課題解決後の自立型ロボットの有用性	7
1.2	本研究の目的	8
1.3	本論文における研究 3 要素	9
1.3.1	研究の学術性	9
1.3.2	研究の新規性	9
1.3.3	研究の有用性	9
1.4	本論文の構成	10

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。それらのロボットというのは、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、「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、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、災害現場での搜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。

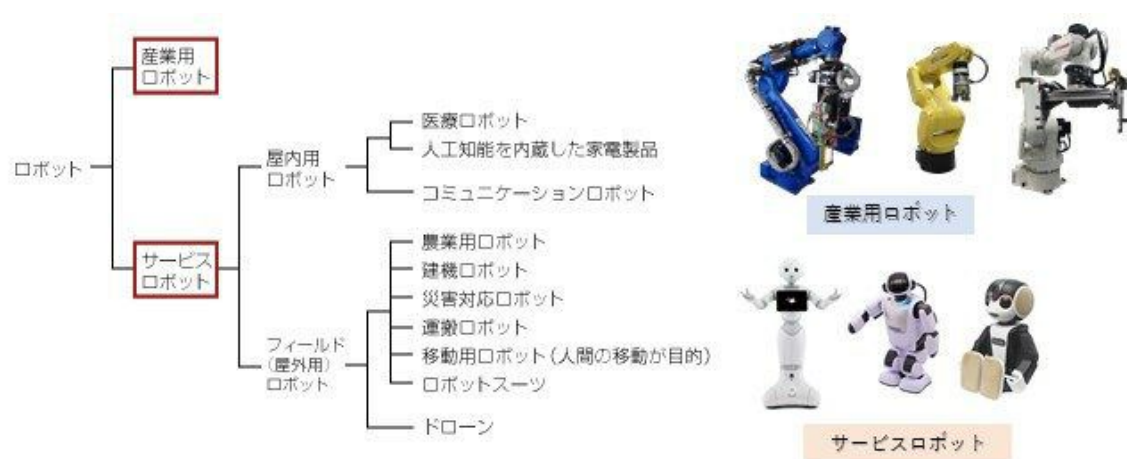


Fig. 1.1: ロボットの分類

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「自動運転に係る制度整備大綱」に下記の図 1.2 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もを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。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、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. 1.2: 自動運転レベルの定義

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。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、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、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リスクを高めているが、いず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] これらの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。

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さらなる研究と開発の必要性を強調している。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潜在能力を持っている。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であろう。

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、災害時の搜索救助活動から、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、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。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。

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します。これまでの研究では、静的障害物の回避は実現されていましたが、動的な障害物の予測と回避は未解決の課題でした。本研究では、知識選択型転移強化学習の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実現し、学習速度の向上や新しい環境への適応度を高めることを目指します。

動的障害物に対する、知識選択型転移強化学習 SAP-net を用いたアプローチ n の検討
とハイパーパラメータの定量的最適化手法の提案

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ます。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ります??。

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ります。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します??。

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ります。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ます。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ります??。

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. 1.3 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、第 1 章の内容を踏まえて本研究のアプローチについて述べる。各章の説明はダイジェストとすべきである。第 3 章では、第 2 章で述べた本研究における強化学習の構成について詳細に述べる。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。

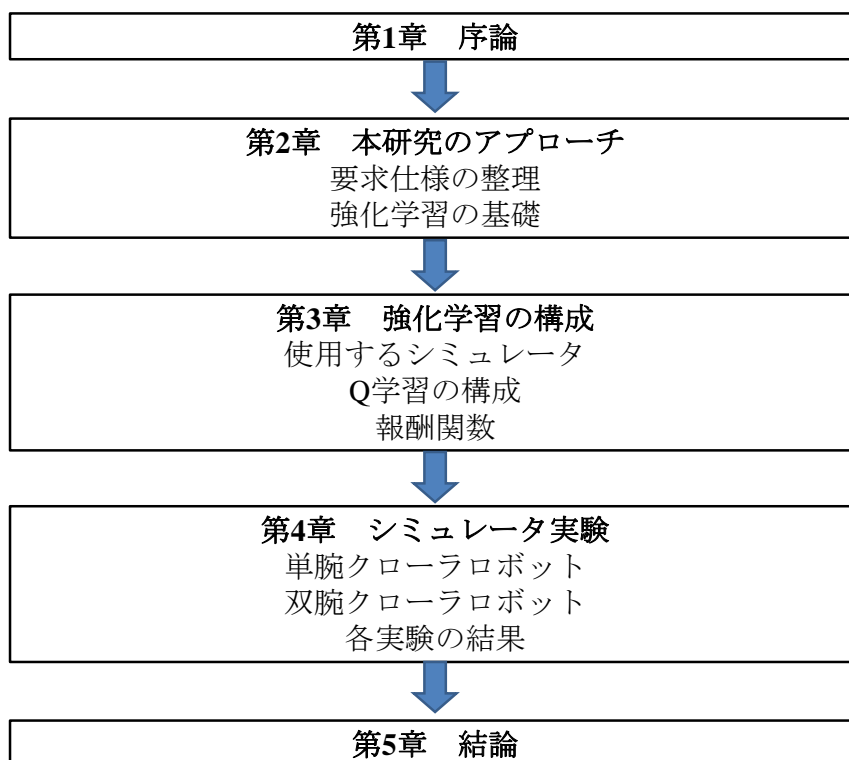


Fig. 1.3: The structure of this thesis

第 2 章

関連研究

Contents

2.1	はじめに	12
2.2	関連研究	13
	2.2.1 強化学習を用いたロボットの行動獲得に関する既存研究	13
2.3	おわりに	14

2.1 はじめに

第 2 章からは、最初のセクションとして「はじめに」をいれる。すなわち、各章の最後のセクションは「おわりに」である。このイントロ的な部分では、本章のあらましを以下のように端的に表現する。

??節では、第 1 章を踏まえて、リカバリモーションを獲得する手法の要求仕様を整理する。??節では、リカバリモーション獲得の方針について述べる。2.2 節では、強化学習を用いたロボットの行動獲得についての関連研究について述べる。最後に 2.3 節で本章のまとめを述べる。

2.2 関連研究

2.2.1 強化学習を用いたロボットの行動獲得に関する既存研究

強化学習はロボットなどの学習の主体が自律的に行動を獲得するために用いる手法であり，多くの研究がなされている [Kober 2013]. 最近では，強化学習と深層学習を組み合わせた手法により，Atari2600 というビデオゲームに対して非常に良い結果を得ており，強化学習は幅広く活用されている [Mnih 2013].

ロボットの移動モーションの獲得にもよく強化学習が用いられる．関連研究として六脚ロボットにおける強化学習を用いた故障時の移動法獲得 [新堀 2009] や強化学習と CPG (Central Pattern Generator) を用いた多脚ロボットの歩行動作獲得の研究 [石倉 2016] が挙げられる．以上の文献からもロボットの行動獲得に強化学習が有用であることがわかる．本節では，以上のロボットの行動獲得について詳しく述べる．

2.3 おわりに

本章では，提案するリカバリモーション獲得のための手法の全体像を述べた．

??節では，第1章を踏まえて，必要なシステムについて整理した．

??節では，それぞれについて強化学習を用いたリカバリモーション獲得をシミュレータ上で行うことを述べた．

2.2節では，強化学習を用いたロボットの行動獲得に関する関連研究について述べた．

??節では，本研究のアプローチを述べ，強化学習を用いた既存研究との関連性について述べた．

??節では，強化学習の基礎的な内容について述べた．強化学習は報酬の設計と価値関数の更新アルゴリズムが重要になることを述べた．

次章では，強化学習を用いたリカバリモーション獲得について詳細に述べる．

Table 2.1: Expectation of fire departments to robotic systems [田所 2012]

災害と必要な機能	期待割合 (%)
CBRN 災害	
センサによる CBRN 物質の特定	80
安全な場所への被災者の搬送	61
CBRN 物質の除去	49
火災	
ビル内の消火	61
ビル内での探索	51
輻射熱にかかわらず消火	49
地震災害	
瓦礫の上からの搜索	53
瓦礫の中での搜索	45
重量瓦礫の除去	43
水害	
要救助者の搜索	55
水の中からの救助	49

第 3 章

提案手法名をここに書く

Contents

3.1	はじめに	16
3.2	本研究のアプローチ	17
3.3	提案手法に用いる技術	17
3.3.1	強化学習の概要とその応用	17
3.3.2	転移学習の概要とその応用	18
3.3.3	転移強化学習の概要	18
3.4	提案システムの構成	19
3.4.1	動的障害物回避への適用	19
3.4.2	知識選択手法の概念	19
3.4.3	知識選択型転移強化学習の原理	19
3.4.4	SAP-net の概念	19
3.5	SAP-net の構成	20
3.5.1	SAP-net の基本構成	20
3.5.2	SAP-net の内部関数	20
3.5.3	SAP-net のプログラム要件	20
3.5.4	SAP-net の速度上限	20
3.5.5	SAP-net の使用方法	20
3.6	数値的な評価指標	20
3.6.1	減衰関数について	20
3.7	最適化シミュレーション	20
3.8	提案システムの全体像	20
3.9	おわりに	21

3.1 はじめに

ここでは，論文を執筆するにあたり，マナーやルールを書いていく．

3.2 本研究のアプローチ

3.3 提案手法に用いる技術

3.3.1 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。

??節で述べたように、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によって、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。

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。

3.3.2 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。

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。

3.3.3 転移強化学習の概要

転移強化学習という技術について説明する。

3.4 提案システムの構成

実際の移動ロボットにおける動的障害物回避アルゴリズムの設計と実装についての説明

3.4.1 動的障害物回避への適用

転移強化学習の動的障害物回避問題への応用方法について説明する。

3.4.2 知識選択手法の概念

知識選択の重要性と、SAP-net の役割と機能について詳細に述べる。

3.4.3 知識選択型転移強化学習の原理

知識選択型転移強化学習の原理提案するアプローチの理論的背景と原理を説明

3.4.4 SAP-net の概念

SAP-net のアーキテクチャ、アルゴリズムの詳細、および実装の工夫について書く

3.5 SAP-net の構成

3.5.1 SAP-net の基本構成

3.5.2 SAP-net の内部関数

3.5.3 SAP-net のプログラム要件

3.5.4 SAP-net の速度上限

3.5.5 SAP-net の使用方法

3.6 数値的な評価指標

3.6.1 減衰関数について

3.7 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。

3.8 提案システムの全体像

パーツの構成全体を見せる

3.9 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	25
4.2	実験環境について	26
4.3	予備実験	27
4.3.1	予備実験目的	27
4.3.2	予備実験条件	27
4.3.3	予備実験結果	27
4.4	静的障害物回避実験	28
4.4.1	静的障害物回避実験目的	28
4.4.2	静的障害物回避実験条件	28
4.4.3	静的障害物回避実験評価	28
4.4.4	静的障害物回避実験結果	28
4.5	動的障害物回避実験	29
4.5.1	動的障害物回避実験目的	29
4.5.2	動的障害物回避実験条件	29
4.5.3	動的障害物回避実験評価	29
4.5.4	動的障害物回避実験結果	29
4.6	知識選択実験	30
4.6.1	動的障害物回避実験目的	30
4.6.2	動的障害物回避実験条件	30
4.6.3	動的障害物回避実験評価	30
4.6.4	動的障害物回避実験結果	30
4.7	最適化実験	31
4.7.1	最適化実験目的	31
4.7.2	最適化実験条件	31
4.7.3	最適化実験結果	31
4.8	実験結果	31
4.9	手法比較	31

4.10	おわりに	32
------	------------	----

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。

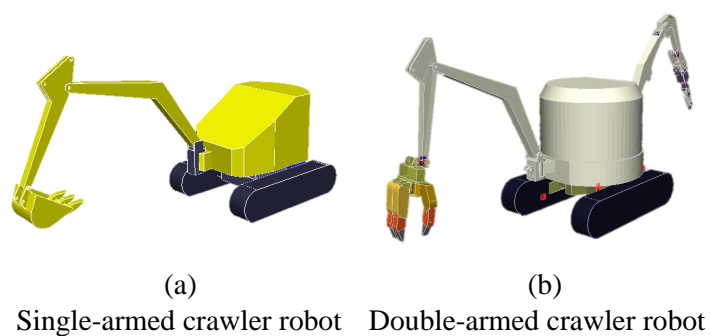


Fig. 4.1: Single-armed and double-armed crawler robot

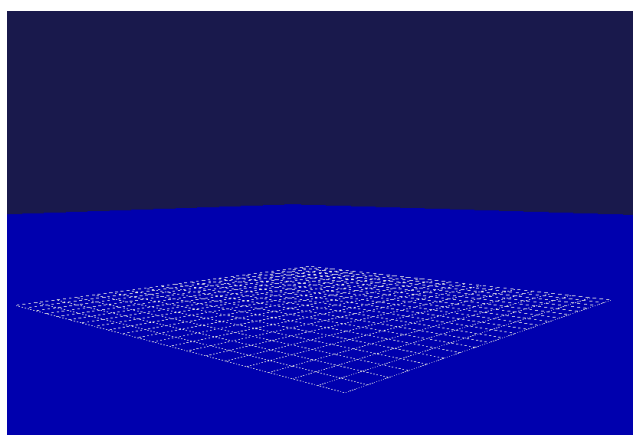


Fig. 4.2: Flat surface in Choreonoid 1.5

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	34
5.2	今後の展望	35

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。

謝辭

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 浅間 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浅間・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 恵さん, 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.

[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.

[浅間 2014]

浅間 一：“災害時に活用可能なロボット技術の研究開発と運用システムの構築,” 日本ロボット学会誌, vol. 32, no. 1, pp. 37–41, 2014.

[田所 2012]

田所 諭：“災害対応ロボットのあるべき姿,” 建設の施工企画, vol. 753, pp. 23–31, 2012.

[田所 2014]

田所 諭：“防災ロボットについて我が国が取り組むべき中長期的課題,” 日本ロボット学会誌, vol. 32, no. 2, pp. 154–161, 2014.

[久武 2011]

久武 経夫, 中里 邦子：“大災害に立ち向かうロボットの開発,” 建設の施工企画, vol. 740, pp. 42–47, 2011.

[江川 2014]

江川 栄治：“小型双腕住器型ロボット「ASTACO-SoRa」,” 日本機械学会誌, vol. 117, no. 1151, pp. 682–683, 2014.

[新堀 2009]

新堀 航太, 兵頭 和幸, 砂山 享祐, 三上 貞芳: “強化学習を用いたモジュール型多脚ロボットにおける適応的移動法獲得,” 情報処理学会論文誌, vol. 50, no. 3, pp. 1170–1180, 2009.

[石倉 2016]

石倉 裕貴, 岸本 良一, 堀内 匡: “CPG と強化学習を用いた多脚ロボットの目標到達行動の獲得,” 電気学会論文誌 C, vol. 136, no. 3, pp. 333–339, 2016.

[Sutton 1998]

R. S. Sutton, A. Gbarto (三上貞芳, 皆川雅章訳) : 強化学習, 森北出版, 2000.

[木村 1999]

木村 元, 宮崎 和光, 小林 重信: “強化学習システムの設計指針,” 計測と制御, vol. 38, no. 10, pp. 1–6, 1999.

[広瀬 1998]

広瀬 茂男, 塚越 秀行, 米田 完: “不整地における歩行機械の静的安定性評価基準,” 日本ロボット学会誌, vol. 16, no. 8, pp. 1076–1082, 1998.

[米田 1996]

米田 完, 広瀬 茂男: “歩行機械の転倒安定性,” 日本ロボット学会誌, vol. 14, no. 4, pp. 517–522, 1996.

[中岡 2013]

中岡 慎一郎: “拡張可能なロボット用統合 GUI 環境 Choreonoid,” 日本ロボット学会誌, vol. 31, no. 3, pp. 12–17, 2013.

[中村 2016]

中村 晋也, 吉灘 裕, 倉鋪 圭太, 谷本 貴頌, 近藤 大祐: “複合ロボットのための動力学シミュレータの開発,” 日本機械学会ロボティクス・メカトロニクス講演会’16 講演論文集, 2016.

[内閣府 2016]

内閣府: “平成 28 年版防災白書,” http://www.bousai.go.jp/kaigirep/hakusho/h28/honbun/3b_6s_10_00.html, 2016, 閲覧日 2016.12.15.

[気象庁 2016]

気象庁: “日本付近で発生した主な被害地震（平成 8 年以降）,” <http://www.data.jma.go.jp/svd/eqev/data/higai/higai1996-new.html#higai2016>, 2016, 閲覧日 2016.12.15.

[原木 2013]

原木俊彦, 大川一也, 加藤秀雄, 樋口静一: “軽微な故障を抱えた脚型ロボットにおけるゴール到達のための自律動作修正,” 日本機械学会ロボティクス・メカトロニクス講演会’13 講演論文集, 2013.

[防衛省 自衛隊 2016]

防衛省・自衛隊: “平成 28 年熊本地震に係る災害派遣について（最終報）,” <http://www.mod.go.jp/j/press/news/2016/05/30b.html>, 2016, 閲覧日 2016.12.16.

[警察庁 2011]

警察庁: “平成 23 年版警察白書,” <https://www.npa.go.jp/hakusyo/h23/honbun/index.html>, 2011, 閲覧日 2016.12.16.

[東京電力 2016]

東京電力ホールディングス: “ロボット技術の活用,” <http://www.tepco.co.jp/decommision/principles/technology/robot/index-j.html>, 2016, 閲覧日 2017.1.4.

<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 2013 Deep 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, 浅間 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ウム予稿集, 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.