

2023 年度 卒業論文



知識選択型転移強化学習を用いた自律型移動ロボットにおける障害物回避

Obstacle Avoidance in Autonomous Mobile Robots Using
Knowledge-Selective Transfer Reinforcement Learning

指導教員 準教授 河野 仁

東京電機大学 工学部 情報通信工学科

学籍番号 20EC060

須賀 哉斗

目次

第1章	序論	1
1.1	背景	2
1.2	自動運転の事故事例	3
1.2.1	自動運転車による死亡事故	3
1.2.2	自動運転レベル4の運行を行っていた車両の接触事故	4
1.3	関連研究	5
1.3.1	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	5
1.3.2	強化学習による仮想環境と実環境における自動走行車いすの障害物 回避	6
1.4	本研究の目的	7
1.5	本論文における研究3要素	8
1.5.1	研究の学術性	8
1.5.2	研究の新規性	8
1.5.3	研究の有用性	8
1.6	本論文の構成	9
第2章	学習アルゴリズム	11
2.1	はじめに	12
2.2	学習アルゴリズム	12
2.2.1	強化学習	12
2.2.2	転移学習	13
2.2.3	行動選択	14
2.2.4	知識選択	14
2.3	おわりに	16
第3章	提案手法	17

3.1	はじめに	18
3.2	提案手法	18
3.2.1	Q学習による強化学習実験	18
3.2.2	SAP-net の活用	18
3.2.3	物理演算シミュレーション	19
3.2.4	実機ロボット	20
3.3	おわりに	21
第4章 実験		23
4.1	はじめに	24
4.2	障害物回避の強化学習実験	25
4.2.1	目的と実験条件	25
4.2.2	実験結果	26
4.3	SAP-net を実装したシミュレーション実験	44
4.3.1	目的と実験条件	44
4.3.2	実験結果	44
4.4	実機実装と障害物回避性能の検証	47
4.4.1	目的と実験条件	47
4.4.2	実験結果	47
4.5	おわりに	51
第5章 結論		53
5.1	結論	54
5.2	今後の展望	55
謝辞		58
参考文献		59
研究業績		62

図目次

1.1	自動運転のレベル分け [ROHM 2020]	2
1.2	事故を起こした Uber Technologies の自動運転車 [米国家運輸安全委員会 2019]	3
1.3	自動運転レベル4の運行を行っていた車両 [ソリトンシステムズ 2023]	4
1.4	電動四輪車の知的自動運転システム [中川 2005].....	5
1.5	電動車いす（左）と3Dモデル（右） [坂田 2022]	6
1.6	本論文の構成	9
2.1	強化学習の簡略図	12
2.2	転移学習の簡略図	13
2.3	SAP-net のイメージ図	15
3.1	Webots の操作画面	19
3.2	シミュレーションで用いるロボットモデル	19
4.1	強化学習シミュレーションの環境	25
4.2	行動回数の推移を表した学習曲線（1番）	26
4.3	行動回数の推移を表した学習曲線（2番）	27
4.4	行動回数の推移を表した学習曲線（3番）	27
4.5	行動回数の推移を表した学習曲線（4番）	28
4.6	行動回数の推移を表した学習曲線（5番）	28
4.7	行動回数の推移を表した学習曲線（6番）	29
4.8	行動回数の推移を表した学習曲線（7番）	29
4.9	行動回数の推移を表した学習曲線（8番）	30
4.10	行動回数の推移を表した学習曲線（9番）	30
4.11	行動回数の推移を表した学習曲線（10番）	31

4.12	獲得報酬の推移を表した学習曲線（1番）	31
4.13	獲得報酬の推移を表した学習曲線（2番）	32
4.14	行動回数の推移を表した学習曲線（3番）	32
4.15	行動回数の推移を表した学習曲線（4番）	33
4.16	行動回数の推移を表した学習曲線（5番）	33
4.17	獲得報酬の推移を表した学習曲線（6番）	34
4.18	獲得報酬の推移を表した学習曲線（7番）	34
4.19	行動回数の推移を表した学習曲線（8番）	35
4.20	行動回数の推移を表した学習曲線（9番）	35
4.21	行動回数の推移を表した学習曲線（10番）	36
4.22	強化学習後の移動軌跡（1番）	37
4.23	強化学習後の移動軌跡（2番）	38
4.24	強化学習後の移動軌跡（3番）	38
4.25	強化学習後の移動軌跡（4番）	39
4.26	強化学習後の移動軌跡（5番）	39
4.27	強化学習後の移動軌跡（6番）	40
4.28	強化学習後の移動軌跡（7番）	40
4.29	強化学習後の移動軌跡（8番）	41
4.30	強化学習後の移動軌跡（9番）	41
4.31	強化学習後の移動軌跡（10番）	42
4.32	SAP-net の知識（方策）ネットワークのイメージ図	43
4.33	行動回数の推移を表した学習曲線	45
4.34	獲得報酬の推移を表した学習曲線	45
4.35	障害物を2つ配置した場合のシミュレーション環境	46
4.36	障害物を2つ配置した場合の移動軌跡	46
4.37	ラズベリーパイマウス	47
4.38	実験環境	48
4.39	障害物を右に回避した時の比較動合成画像	49
4.40	障害物を右に回避した時の比較動合成画像	49
4.41	障害物を右に回避した時の知識選択の推移	50
4.42	障害物を左に回避した時の知識選択の推移	50

表目次

4.1	強化学習の条件および学習パラメータ	25
-----	-------------------------	----

第1章

序論

Contents

1.1	背景	2
1.2	自動運転の事故事例	3
1.2.1	自動運転車による死亡事故	3
1.2.2	自動運転レベル4の運行を行っていた車両の接触事故	4
1.3	関連研究	5
1.3.1	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 ...	5
1.3.2	強化学習による仮想環境と実環境における自動走行車いすの障 害物回避	6
1.4	本研究の目的	7
1.5	本論文における研究3要素	8
1.5.1	研究の学術性	8
1.5.2	研究の新規性	8
1.5.3	研究の有用性	8
1.6	本論文の構成	9

1.1 背景

自動車業界は AI や IoT のような最先端技術によって「100 年に一度の大変革の時代」に突入しているとされ、その中心には自動運転技術がある。特に、2023 年 4 月 1 日には日本で自動運転レベル 4 の公道走行が解禁された。自動運転レベル 4 とは、限定された走行領域でシステムが全ての運転操作を実施し、ドライバーが運転席を離れることができる段階である。自動運転のレベル分けを Fig. 1.1 に示す。2025 年を目指して完全自動運転、レベル 5 の実現が目指されている。レベル 5 ではどのような運転環境下でも、人間の介入なしに運転が可能となる。自動運転技術の実用化により、安全性の向上や、運送効率の向上、新たな交通サービスの創出等が図られて、大幅な生産性向上に資する可能性を秘めている [国土交通省 2016]。他にも渋滞の解消や緩和、技術・ノウハウに基づく国際展開など、自動運転の実現により期待される効果があるが、最も期待されているのが交通事故の低減である。2022 年の交通事故死者数は 3541 人で、年間一日当たり 9.7 人が亡くなっている [厚生労働省 2023]。悲惨な交通事故のほとんどはわき見運転や安全運転義務違反など、運転者のミスに起因しているが、自動運転の実現により運転者のミスに起因する事故の防止に効果があるといわれている。自動運転の実現に向けて、日本を含めた世界各国では、自動運転技術が搭載された車両実験が進められている。このような実証実験は、自動運転技術の安全性や信頼性を検証するために不可欠となる。しかし、自動運転技術に発展と並行して、自動運転車に関する事故も多く発生している。

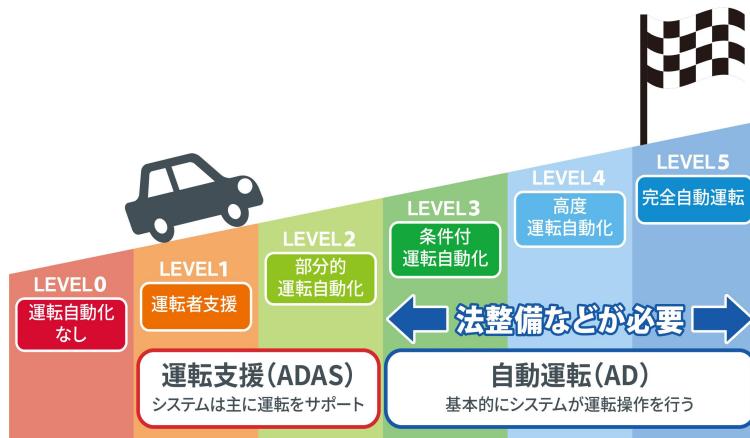


Fig. 1.1: 自動運転のレベル分け [ROHM 2020]

1.2 自動運転の事故事例

1.2.1 自動運転車による死亡事故

2018年3月、米ライドシェア大手の Uber Technologies（ウーバー・テクノロジーズ）の自動運転車が米アリゾナ州を走行中、自転車を押しながら道路を横断していた歩行者をはね、死亡させる事故を起こした。事故を起こした車両を Fig.1.2 に示す。当時の米の自動運転のレベルは 3 である。自動運転レベル 3 は、限定された条件下において、システムが全ての運転操作を実施する。ただし運転自動化システム作動中であっても、システムからの要請があればドライバーはいつでも運転に戻れる状態である必要があるレベルである。事故の主要因として、同乗していたセーフティドライバーが事故発生時に携帯端末で動画を視聴していたことが挙げられる。また、ウーバーの自動運転システムが歩行者を検知できず、「自動車」や「自転車」、「その他のオブジェクト」などと認識していたことも発覚している。その挙動についても「左側の車線を走行」「静止中」などと判断したため、直前まで衝突の危険性を検知できなかった。

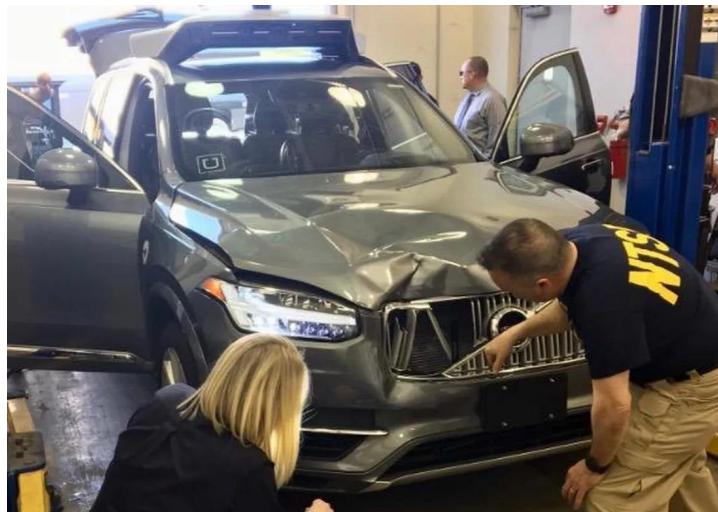


Fig. 1.2: 事故を起こした Uber Technologies の自動運転車 [米国家運輸安全委員会 2019]

1.2.2 自動運転レベル4の運行を行っていた車両の接触事故

2023年10月29日には自動運転レベル4の運行を行っていた車両が自転車に接触するという事故が発生した。この車両は自動運転レベル4の分類される高度な自動化を実現しており、障害物回避の検知して避けるためのセンサやレーダが搭載されていた。この事故が起きた主な原因は、車両の自動運転ブレーキが自転車を適切に認識せず、期待通りに作動しなかったことにある。自動運転システムが自転車を認識しなかった背景には、学習データの不測が大きく関係していることが判明した。自動運転車両の学習アルゴリズムは、様々な状況下データを基にして障害物を認識して適切な行動をするように訓練される。しかしこの車両の場合、特定の状況下での自転車との遭遇に関するデータが学習過程で不足していたため、システムが自転車を認識したが、適切な安全措置をとることができなかった。



Fig. 1.3: 自動運転レベル4の運行を行っていた車両 [ソリトンシステムズ 2023]

1.3 関連研究

1.3.1 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

これまでの研究では、運転技術や運転に関する知識が不十分なユーザにも安心して使用できるように、電動四輪車に人間の運転知識を組み込んだ自動運転システムの開発に注力した [中川 2005]. その電動四輪車を Fig1.4 に示す. この自動運転システムは、特に複雑な交通環境や予測が困難な歩道上での動的障害物に対応する能力に焦点を当て、予測的ファジー制御技術を活用している。この技術により、システムはリアルタイムで環境を解析し、即時かつ適切な運転判断を下すことが可能になる。実際の道路環境を模倣したシミュレーション実験を通じて、この自動運転システムの効果を検証した。実験結果は、特に予期しない障害物に対する回避行動の改善において、このシステムの有効性を示している。しかし、実機への実装はまだ行われておらず、今後の課題として期待されている。



Fig. 1.4: 電動四輪車の知的自動運転システム [中川 2005]

1.3.2 強化学習による仮想環境と実環境における自動走行車いすの障害物回避

Unity 3D を用いてシミュレーション環境を作成し、そこで車いすの 3D モデルを使用して障害物回避の強化学習を行う [坂田 2022]. 学習アルゴリズムは Deep Deterministic Policy Gradient (DDPG) を採用している. この手法は、行動の選択とその行動に対する価値を別々のニューラルネットワークで学習することにより、連続的な行動空間を扱うことが可能となる. その後、学習済みのモデルを実機である WHILL 社製の電動車いすに適用して、実環境での障害物回避能力を評価している. 障害物が一定距離内にあると認識した場合に自動走行から障害物回避への切り替えを行うルールベースのアプローチで行っているが、新しい状況や環境に適応するためには、手動での調整や追加のプログラミングが必要となってしまう.

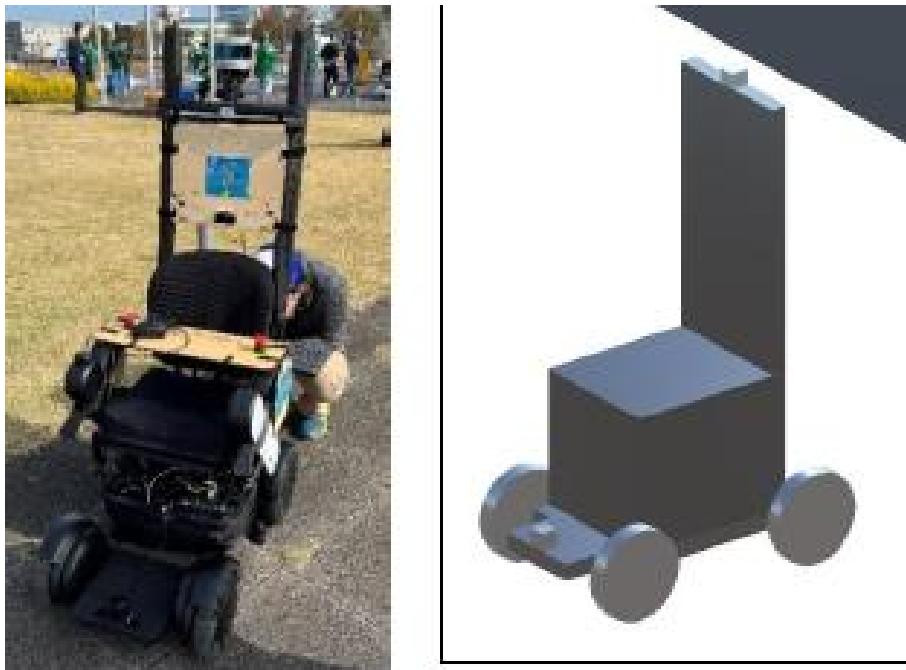


Fig. 1.5: 電動車いす（左）と 3D モデル（右）[坂田 2022]

1.4 本研究の目的

1.1 節より自動運転システムは日々進化しているが、1.2 節のような自動運転にまつわる事故が多発している。このことから、自動運転システムに搭載された環境認知をするためのセンサやレーダに頼るだけでは現時点では限界があることがわかる。2025 年を目途に完全自動運転、レベル 5 の実現が目指されているが、そのためには周囲の物体を正確に識別し、適切な判断を下す能力が不可欠である。現在は機械学習技術を用いて状況判断や障害物検知などを行うことで自動運転技術を実現しているが、障害物回避技術に関しては完全な解決には至っていない。現時点では、関連研究の他にこの課題への対応として、単眼カメラと三次元地図を用いて動的な障害物を三次元的に検出し、その形状を復元するための手法や、動的な障害物を回避するための軌道生成法が研究されているが、いずれも実機を用いた障害物の回避は実現されていない [敷島 2021][金原 2022]。このため完全自動運転を実現するためには、環境認知と行動計画を人工知能に任せると共に、さらに人間の賢さに似た知的能力が必要とされている [勞 2018]。障害物回避を含め、様々な配置パターン環境や場面に対応する手法の 1 つとして転移強化学習が注目されている。中でも知識選択型転移強化学習を用いて開発した Kono らの SAP-net に注目する [Kono2019]。SAP-net を用いて物理演算シミュレータで走行経路の強化学習をした後、学習結果をシニアカーに転移させて自律走行の実現に成功している [河野 2022]。そこで、本研究の目的を以下のとおりとする。

知識選択型転移強化学習 (SAP-net) を活用した移動ロボットにおける障害物回避の実現

1.5 本論文における研究3要素

1.5.1 研究の学術性

本研究の学術性は、様々な障害物の配置パターンを回避する強化学習を行い、強化学習で得られた学習データを知識（行動価値関数）として保存し、活性化拡散モデルを元にした SAP-net を用いた知識選択型の転移強化学習を行うところにある。

1.5.2 研究の新規性

本研究の新規性は、SAP-net における行動選択において、候補となる知識数が多いことがある。SAP-net を用いてシニアカーの障害物回避システムを実現しているが、知識数が少ないため複雑な環境においては適応性が低い。本研究では、知識数を増やすことにより、複雑な環境においても適応的に行動することが可能となる。

1.5.3 研究の有用性

研究の有用性は、自動運転車両やその他の自律型移動ロボットの安全性と効率性を向上させることにある。特に、自動運転システムの性能の向上に伴い、交通事故のリスクを低減する。またこの技術は、都市計画や公共交通システムの最適化にも応用可能であり、広範な影響を与えることが可能となる。

1.6 本論文の構成

本論文の構成について Fig.1.6 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、本研究に対する事故事例、自動運転技術の関連研究、本研究の目的について述べた。

第 2 章では、本研究の目的で用いる学習アルゴリズムとして用いる強化学習、転移学習、行動選択、知識選択の基本知識について述べる。

第 3 章では、第 2 章の学習アルゴリズムを用いた本研究の目的を実現するための提案手法について述べる。

第 4 章では、第 3 章で述べた提案手法の有用性を示すシミュレーション実験、および実機を用いた実験の概要、各実験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。

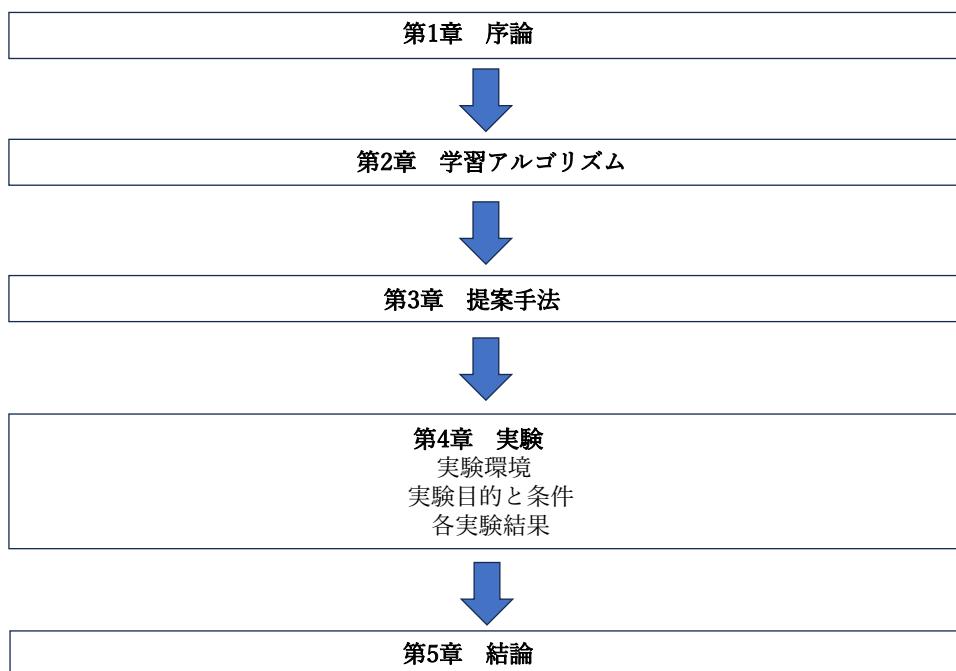


Fig. 1.6: 本論文の構成

第2章

学習アルゴリズム

Contents

2.1	はじめに	12
2.2	学習アルゴリズム	12
2.2.1	強化学習	12
2.2.2	転移学習	13
2.2.3	行動選択	14
2.2.4	知識選択	14
2.3	おわりに	16

2.1 はじめに

本章では本研究で用いる学習アルゴリズムの詳細について述べる。構成は 2.1 節で強化学習，2.2 節で転移学習，2.3 節で行動選択，2.4 節で知識選択手法について述べていく。

2.2 学習アルゴリズム

本研究では、移動ロボットにおける障害物回避のために用いる強化学習，転移学習，行動選択、知識選択を用いる。それぞれ式を示しながら述べていく。

2.2.1 強化学習

タスク達成を目指し、繰り返し最適解を試行錯誤的に学習する手法である。試行錯誤的に学習していくため、知識が全くない状態からの知識の取得が可能である。しかし、試行錯誤的に学習するため、学習が遅いという欠点がある [Sutton1998]。強化学習の流れを Fig.2.1 に示す。

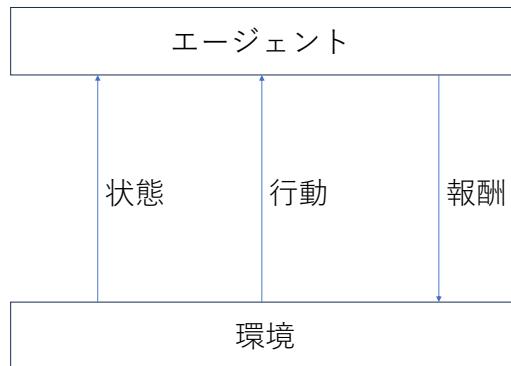


Fig. 2.1: 強化学習の簡略図

強化学習の主体をエージェントと呼ぶ。ある環境において何かしらの行動を起こしたら、その行動から得られる報酬を獲得するという処理を何度も反復することで、報酬の合計が一番大きくなるような行動を学習していく。以下に強化学習に用いられる Q 学習の式を示す。

$$Q(s_t, a) \leftarrow Q(s_t, a) + \alpha \left\{ r + \gamma \max_{a' \in A} Q(s_{t+1}, a') - Q(s_t, a) \right\} \quad (2.1)$$

ここで、 $Q(s_t, a)$ は時刻 t における状態 s から行動 a を選択したときの報酬の期待値を表す行動価値関数を示している。学習率 α ($0 < \alpha \leq 1$) は、更新される価値の大きさおよび学習の速度に影響を与える、高い値では学習が速くなるものの、最適解を見つける確率が低下する可能性がある。割引率 γ ($0 < \gamma \leq 1$) は、将来の報酬の現在価値を計算する際に用いられ、 γ の値が小さい場合は未来の価値を低く評価し、大きい場合は高く評価する。状態 s_{t+1} での最大行動価値を求めるために $\max_{a' \in A} Q(s_{t+1}, a')$ を用い、この値を γ で割引き、現在の状態 $Q(s_t)$ の価値として加算し、さらに報酬 r を加え、 α で割引いた値を $Q(s_t, a)$ に加えることで、行動価値を更新していく。

2.2.2 転移学習

強化学習である程度学習した知識を別のタスクに適用させる手法である。Fig.2.2 に転移学習の簡略図を示す。

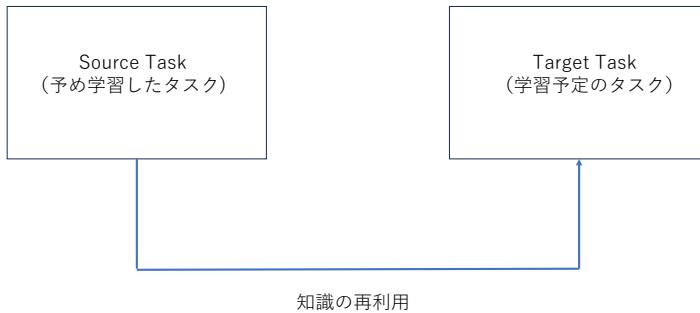


Fig. 2.2: 転移学習の簡略図

Source Task（予め学習したタスク）ではある特定の問題やタスクに対する学習を指す。このタスクにおいてモデルが学習され、特定の知識が蓄積される。Target Task では Source Task とは異なるが、類似性や共通の特徴がある別の問題やタスクを指す。転移学習では Source Task で学習された知識を Target Task で再利用して学習することで、学習の短縮や精度を向上させることが可能である [Taylor2009]。以下に転移学習の式を示す。

$$Q_c(s, a) = Q_c(s, a) + \tau Q_s(s, a) \quad (2.2)$$

$Q_s(s, a)$ は Source task から転移された行動価値関数を示している。 $Q_s(s, a)$ に転移率 τ ($0 < \tau \leq 1$) を掛けることで、新たな環境が再利用される行動価値関数を獲得した環境と異

なる場合においても適応するようになる。 $Q_t(s, a)$ は転移先のタスクで更新する行動価値関数を示している。さらに新たな環境とで学習した行動価値関数も $Q_t(s, a)$ に更新していく。 $Q(s_c, a)$ は転移された行動価値関数と Target Task で獲得した行動価値関数を統合した行動価値関数であり、Target Task で行われる行動選択は $Q_c(s, a)$ を用いて行われる。

2.2.3 行動選択

Q テーブルにはある状態 s で取り得る行動 a とそれに対応する価値が記録されている。価値に基づき、どの行動を取るかを決めるために行動選択関数を用いるアルゴリズムが存在する [河野 2022]。本研究では、行動選択関数ボルツマン選択を用いる。複雑で不確実性の高い環境や、エージェントが幅広い行動から学習によるする必要がある場合に有効である。ボルツマン選択を用いることで、エージェントが受け取る報酬を行動選択に反映させることができる。以下にボルツマン選択の式を示す。

$$p(a|s) = \frac{\exp(\frac{Q(s,a)}{T})}{\sum_{b \in A} \exp(\frac{Q(s,b)}{T})} \quad (2.3)$$

$p(a|s)$ は状態 s において行動選択 a を選択する確率で、 $Q(s, a)$ は行動 a を選択したときの価値、 T は温度定数でボルツマン選択におけるランダム生成を調整するパラメータである。 T が高いほど選択はランダムに近くなり、 T が低いほど最も価値の高い行動が選択されやすくなる。

2.2.4 知識選択

人の脳内における概念の選択手法と言われている活性化拡散モデルを元にした SAP-net (Spreading Activation Policy-network) を用いた知識選択型の転移学習手法を使用し、転移強化学習を行う。使用する知識、いわゆる方策は強化学習で獲得した学習結果を指し、それらを選択することで行動を決定する。SAP-net のイメージ図を Fig.2.3 に示す。

この手法は、強化学習で得られた方策や行動価値関数をグラフ構造で管理し、外部刺激に反応して関連するポリシーの活性値が増加するメカニズムを持っている。この活性値の増加は、グラフ内の他のノードへ伝播し、一定の閾値を超えた方策や行動価値関数が「想起」されて転移学習に利用される。また、SAP-net には時間経過に伴う活性値の減少を考慮する設計が取り入れられており、効率的な情報処理を可能にしている。この手法を式で表したのが活性化拡散方程式である。以下に活性化拡散方程式を示す。

$$A_{j,t+\Delta t} = A_{j,t} + a_\delta + \sum_{k=1}^q \eta_k - D_k \quad (2.4)$$

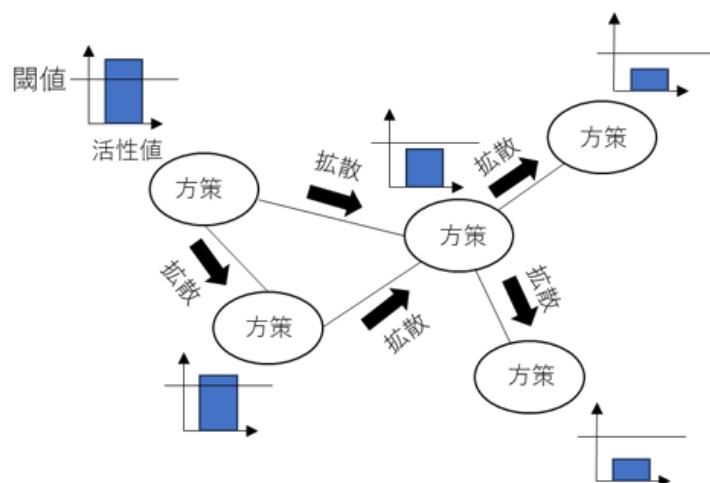


Fig. 2.3: SAP-net のイメージ図

ある時点 t での方策が持つ活性値 $A_{j,t}$ を外部からの入力や内部メカニズムによって加算される a_δ , 周囲の状況や他の方策から拡散されてくる値の総和 $\sum_{k=1}^q \eta_k$ を加えて, さらに時間の経過に伴う自然な減少や外部からの抑制的な影響を反映する減衰定数 D_k を引くことで, 時間 $t + \Delta t$ で活性値 $A_{j,t+\Delta t}$ を更新していく.

2.3 おわりに

本章では知識選択による転移強化学習を用いるために使用する学習アルゴリズムについて述べた。次章では本章で述べた学習アルゴリズムを活用するための提案手法について述べる。

第3章

提案手法

Contents

3.1	はじめに	18
3.2	提案手法	18
3.2.1	Q学習による強化学習実験	18
3.2.2	SAP-net の活用	18
3.2.3	物理演算シミュレーション	19
3.2.4	実機ロボット	20
3.3	おわりに	21

3.1 はじめに

本章では、提案手法の概要、シミュレーション環境について述べる。続く3.2節では、提案手法の詳細について述べる。

3.2 提案手法

本研究の目的は、Q学習を基礎とした強化学習アルゴリズム、新たな知識選択型転移強化学習メカニズムであるSAP-netを組み合わせることにより、ロボットが複雑な障害物環境下での適応と学習をより効率的に行えるようにする新しいアプローチを開発し、その有効性を実験的に検証することにある。このアプローチは、ロボットが未知の障害物環境に迅速に適応、効果的な回避策を学習し、適用する能力を大幅に向上させることを目指している。この目的を達成するために、高度な物理演算を実現するシミュレータWebotsを使用し、ロボットが様々な障害物配置を含む環境下での行動策を効率的に学習できるかどうかを検証する。

3.2.1 Q学習による強化学習実験

初めに複数の配置パターンの障害物配置が含まれる環境下でQ学習による強化学習実験を行う。Q学習は、エージェントが取るべき最適な行動を学習するための一種の価値ベースの強化学習手法であり、各状態における行動の価値（Q値）を推定することにより、最適なポリシーを導出することを目指す。この実験では、ロボットが未知の障害物環境に置かれた際に、自己の位置から目標地点までの経路を最適化する過程を学習する。実験の初期段階では、ロボットはランダムな行動を取ることから始まり、その結果として得られる報酬を基に、徐々に最適な行動方針を学習していく。このプロセスの中心にあるのは、行動の結果として得られる即時報酬と将来の報酬の合計を最大化することによって、最適な行動選択を行うことである。

3.2.2 SAP-netの活用

SAP-netはロボットの初期座標を基準に障害物までの角度と距離を保存させる。この情報はロボットが環境を理解し、障害物を正確に把握するための知識ネットワークを構築する基礎となる。次に強化学習から得た行動価値関数の情報を取得し、それらの類似度を計算する。類似度計算は異なる状況における最適な行動の一貫性を確認する重要なステップである。そして計算された類似度情報を基にネットワークを構築する。このネットワークは活性

化拡散モデルとして機能し、異なる行動価値関数を統合し、環境への柔軟な対応を可能にする。そして構築されたネットワークは知識を選択し、動的な状況において障害物回避の戦略を提供する。これによりロボットは瞬時の判断により、適切な知識を活用して効果的な障害物回避を実現する。

3.2.3 物理演算シミュレーション

強化学習は試行錯誤的に行動を何千と繰り返して学習するため、実環境でやると時間がかかるってしまう。また実機で使用するロボットが破損する恐れがある。そこで Webots という物理演算シミュレーションを使用する [Webots1998]。Webots の操作画面を Fig.3.1 に示す。Webots を使用することで、シミュレーション内のロボットが強化学習を行っていくため、実環境で行うよりも安全に効率よく学習することが可能である。使用するロボットモデルを Fig.3.2 に示す。また障害物の認識は LIDAR を使用する。

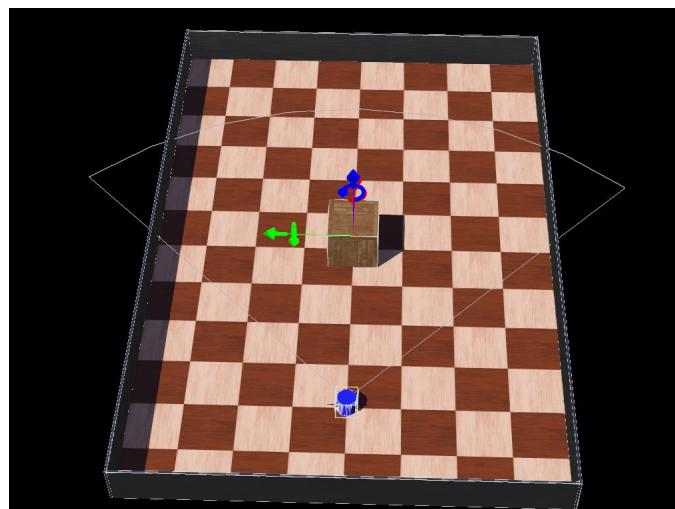


Fig. 3.1: Webots の操作画面



Fig. 3.2: シミュレーションで用いるロボットモデル

3.2.4 実機ロボット

物理演算シミュレータによる強化学習実験および SAP-net を活用したシミュレーション実験を終えた後、強化学習で得られた学習データと SAP-net を USB メモリを用いて実機ロボットに転移し、実機ロボットを用いた実環境での障害物回避を検証する。使用するロボットは Fig.3.2 と同じ二輪ロボットを使用する。また実環境で使用する LIDAR は、webots で使用する LIDAR の測定範囲に合わせる。これにより、シミュレータで開発されたモデルが実環境に適応するのかを検証する。

3.3 おわりに

本章では提案手法について述べた。3.2 節では提案手法として、Q 学習による強化学習実験、SAP-net の活用方法、物理演算シミュレーション、実機ロボットについて述べた。次章では提案手法に基づいた実験について述べる。

第4章

実験

Contents

4.1	はじめに	24
4.2	障害物回避の強化学習実験.....	25
	4.2.1 目的と実験条件	25
	4.2.2 実験結果	26
4.3	SAP-net を実装したシミュレーション実験.....	44
	4.3.1 目的と実験条件	44
	4.3.2 実験結果	44
4.4	実機実装と障害物回避性能の検証	47
	4.4.1 目的と実験条件	47
	4.4.2 実験結果	47
4.5	おわりに	51

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実験について述べる。4.2節ではWebotsで障害物回避の強化学習シミュレーションによる実験の目的、実験条件、および実験結果について述べる。4.3節では、SAP-netを実装したシミュレーション実験の目的、実験条件、および実験結果について述べる。4.4節では学習結果を転移させた実機ロボットによる障害物回避の実験の目的、条件、および結果について述べる。

4.2 障害物回避の強化学習実験

4.2.1 目的と実験条件

事前実験として強化学習を用いて障害物を回避する実験を行った。事前実験の目的は障害物回避の知識を実機に転移させるためである。本実験の条件および学習パラメータを Table.4.1 に示す。実験環境を Fig.4.1 に示す。

Table 4.1: 強化学習の条件および学習パラメータ

エピソード数 (回)	7000
ステップ数	4 (前進・後退・右旋回・左旋回)
強化学習の手法	Q 学習
行動価値の配列数	5 (x 座標, y 座標, ロボットの向き, 障害物までの角度, 障害物までの最小距離)
正の報酬の付与条件	ゴール到達時
負の報酬の付与条件	障害物に衝突またはゴールから遠ざかる行動をした時
学習率	0.1
割引率	0.9
ボルツマン選択の温度定数	0.5

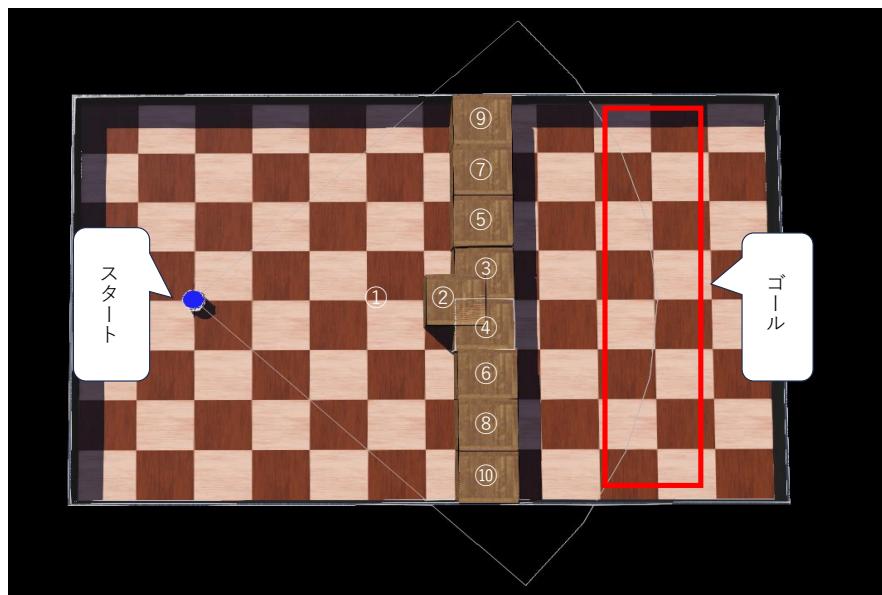


Fig. 4.1: 強化学習シミュレーションの環境

上記の条件・学習パラメータ・環境で強化学習を進め、エピソードごとのステップ数を分

析して、障害物の回避策を効率的に実施するための知識を構築する。

4.2.2 実験結果

以下に Fig.4.1 の配置パターン (1 番から 10 番) の強化学習の結果を示す。Fig.4.2～Fig.4.11 に行動回数の推移を表した学習曲線、Fig.4.12～Fig.4.21 に獲得報酬の推移を表した学習曲線を示す。

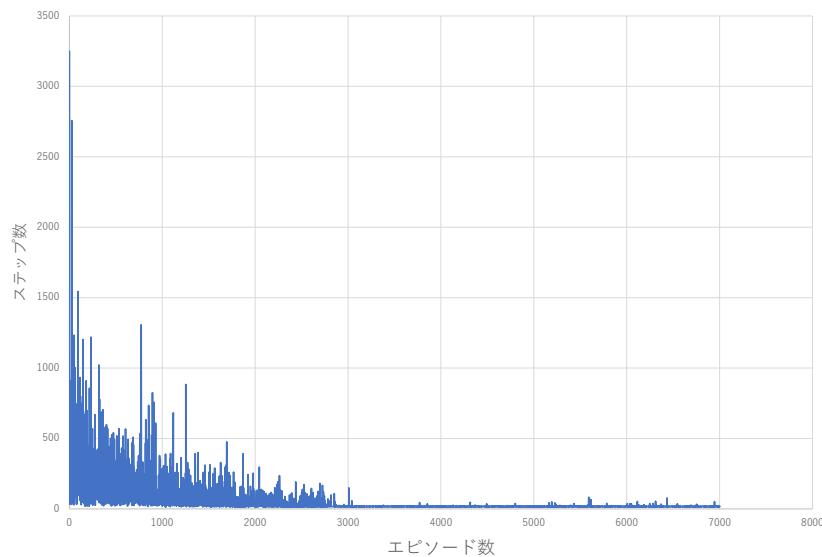


Fig. 4.2: 行動回数の推移を表した学習曲線 (1番)

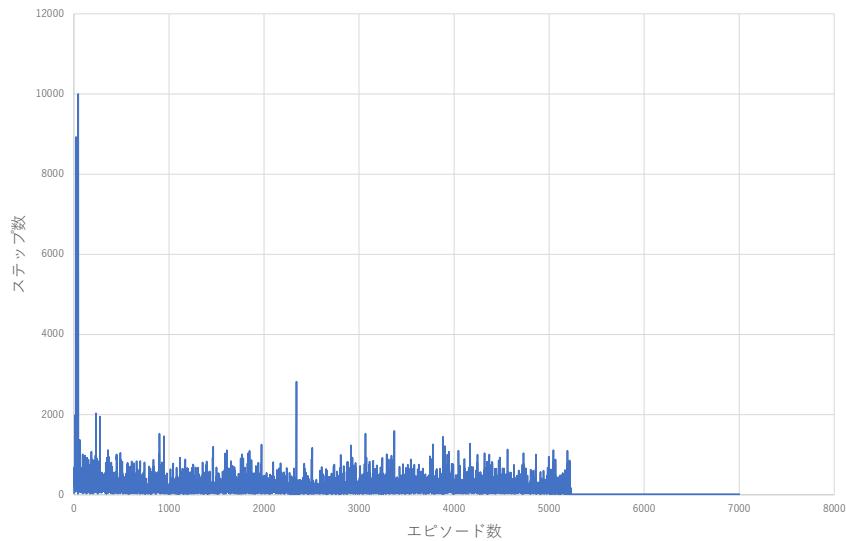


Fig. 4.3: 行動回数の推移を表した学習曲線（2 番）

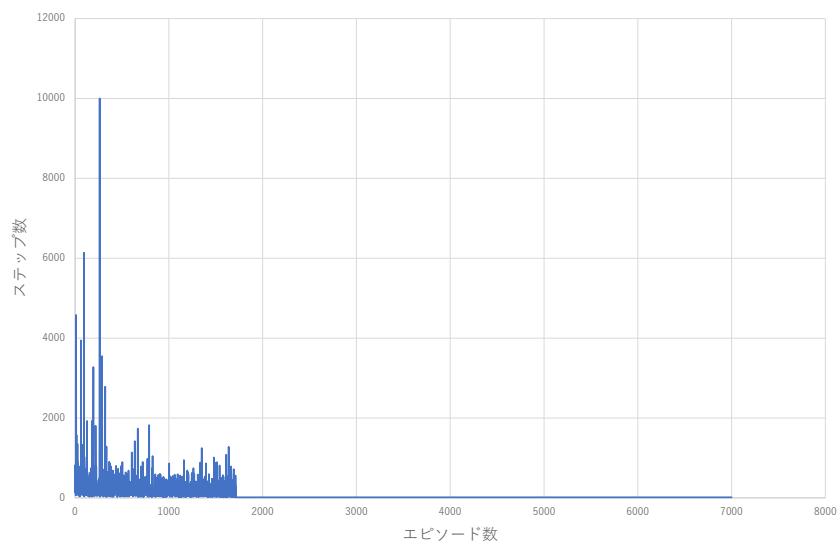


Fig. 4.4: 行動回数の推移を表した学習曲線（3 番）

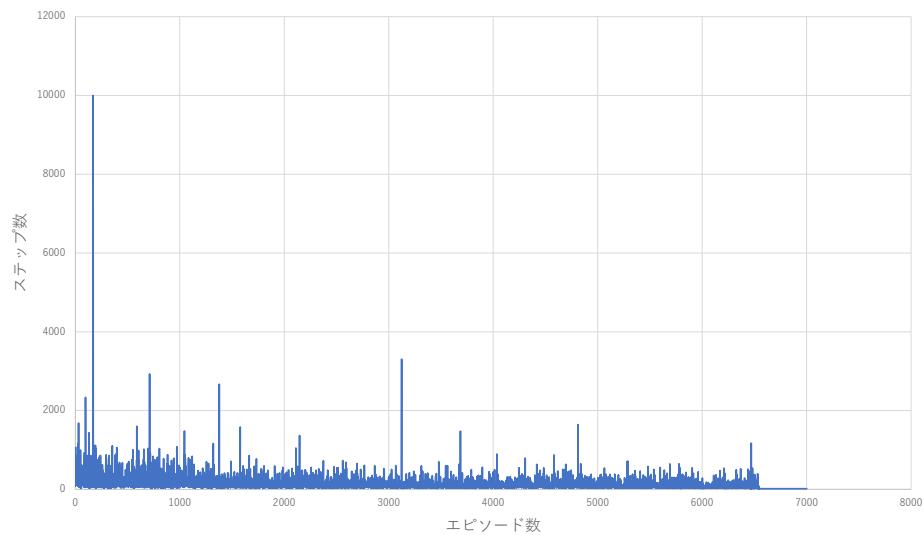


Fig. 4.5: 行動回数の推移を表した学習曲線（4番）

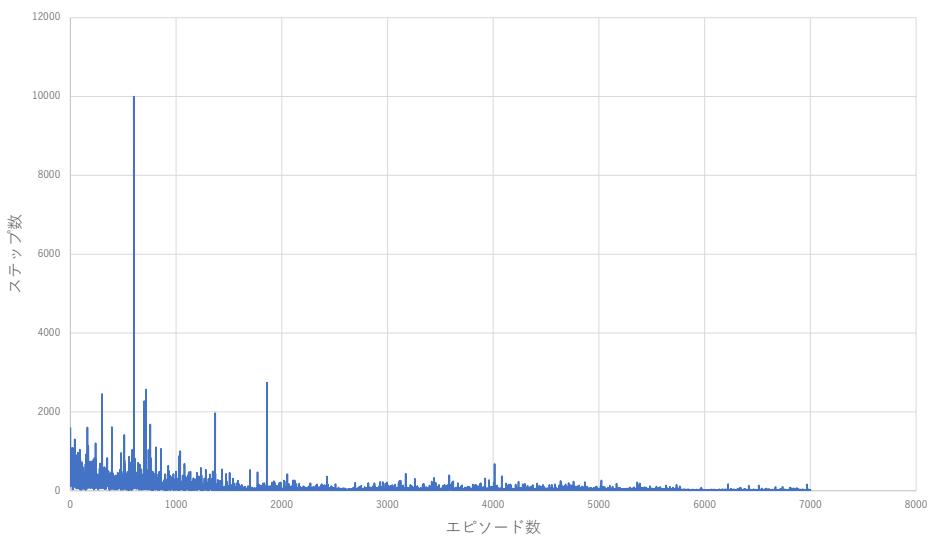


Fig. 4.6: 行動回数の推移を表した学習曲線（5番）

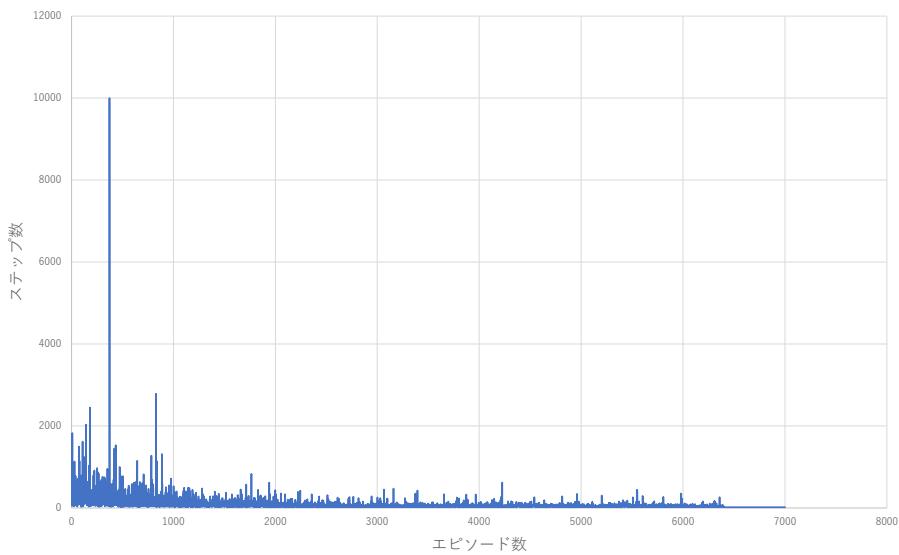


Fig. 4.7: 行動回数の推移を表した学習曲線（6番）

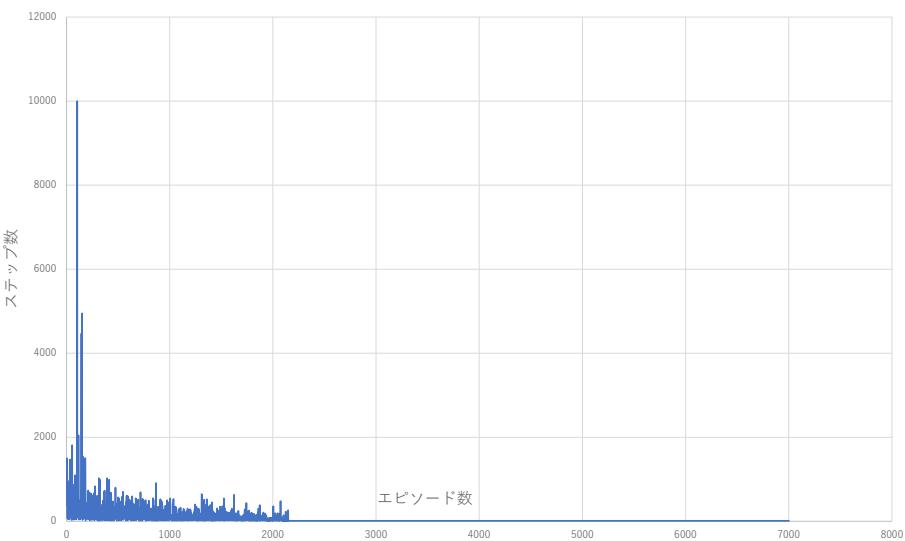


Fig. 4.8: 行動回数の推移を表した学習曲線（7番）

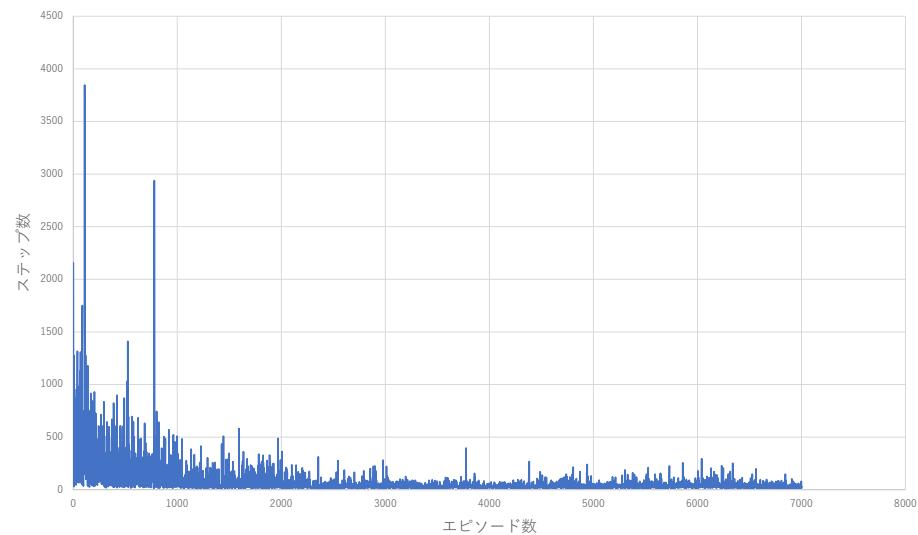


Fig. 4.9: 行動回数の推移を表した学習曲線（8番）

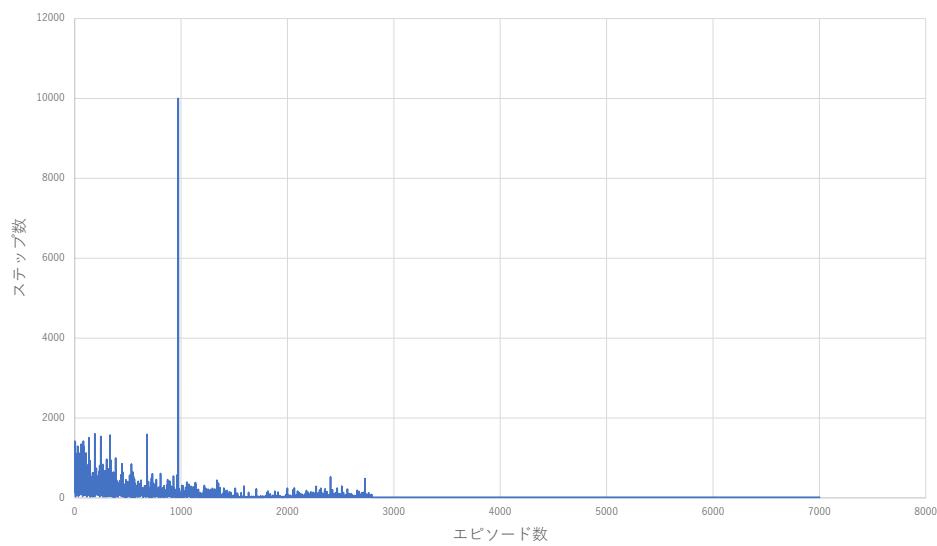


Fig. 4.10: 行動回数の推移を表した学習曲線（9番）

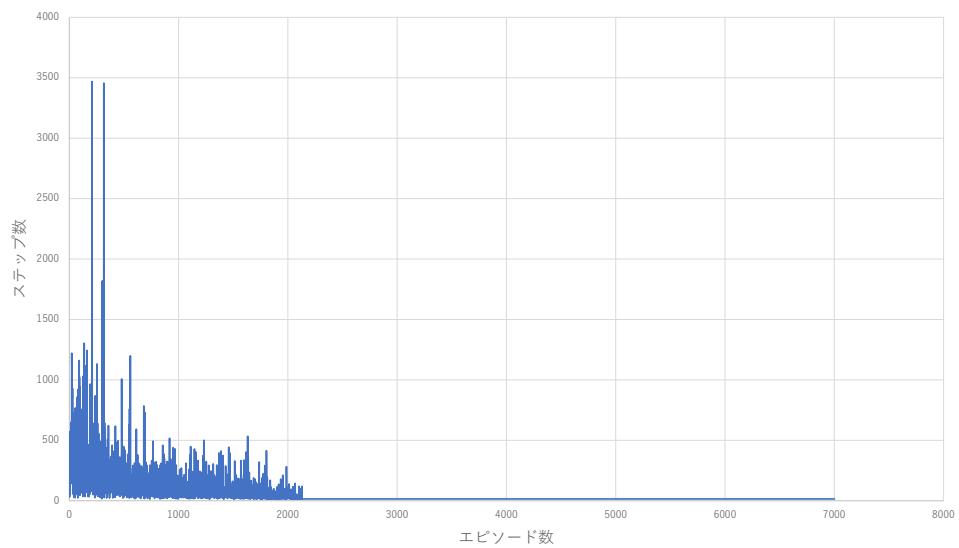


Fig. 4.11: 行動回数の推移を表した学習曲線（10 番）

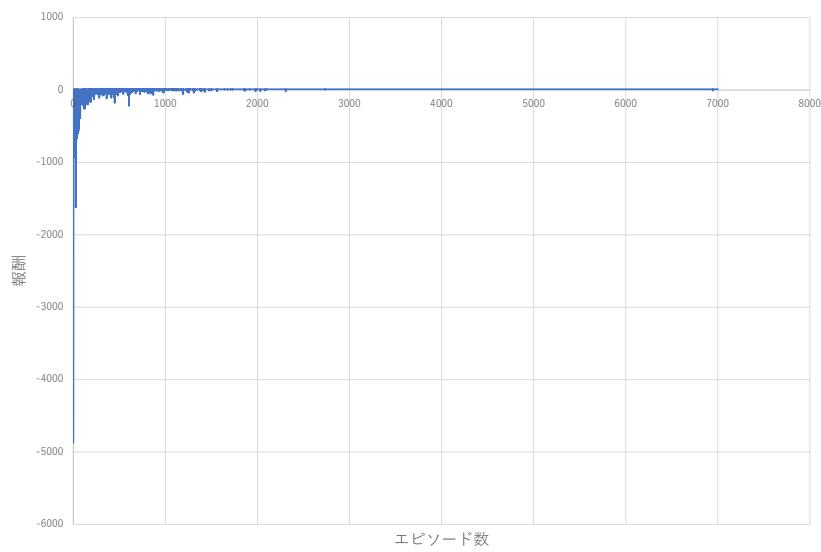


Fig. 4.12: 獲得報酬の推移を表した学習曲線（1 番）

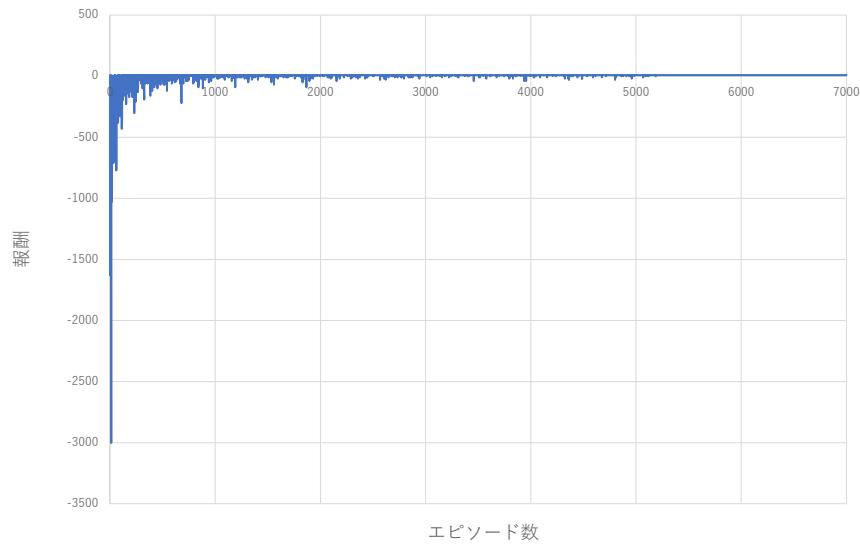


Fig. 4.13: 獲得報酬の推移を表した学習曲線（2 番）

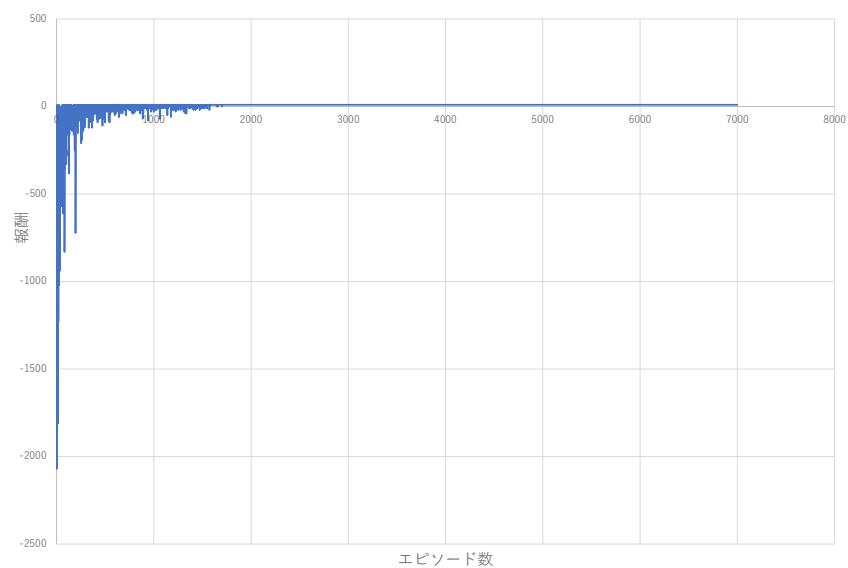


Fig. 4.14: 行動回数の推移を表した学習曲線（3 番）

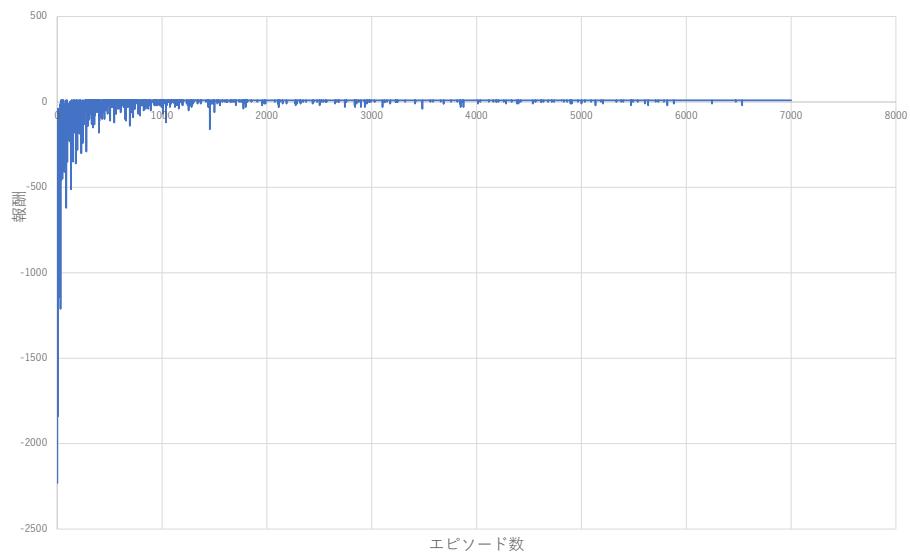


Fig. 4.15: 行動回数の推移を表した学習曲線 (4 番)

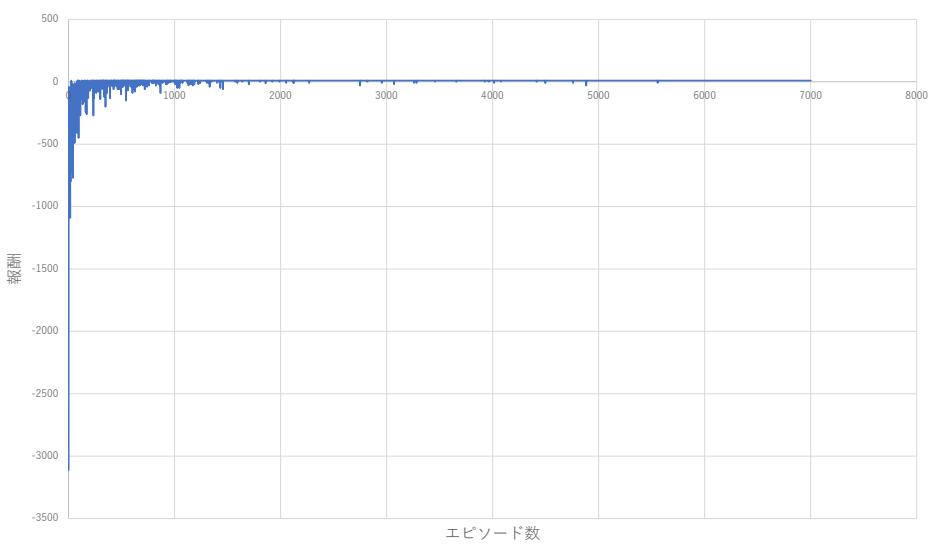


Fig. 4.16: 行動回数の推移を表した学習曲線 (5 番)

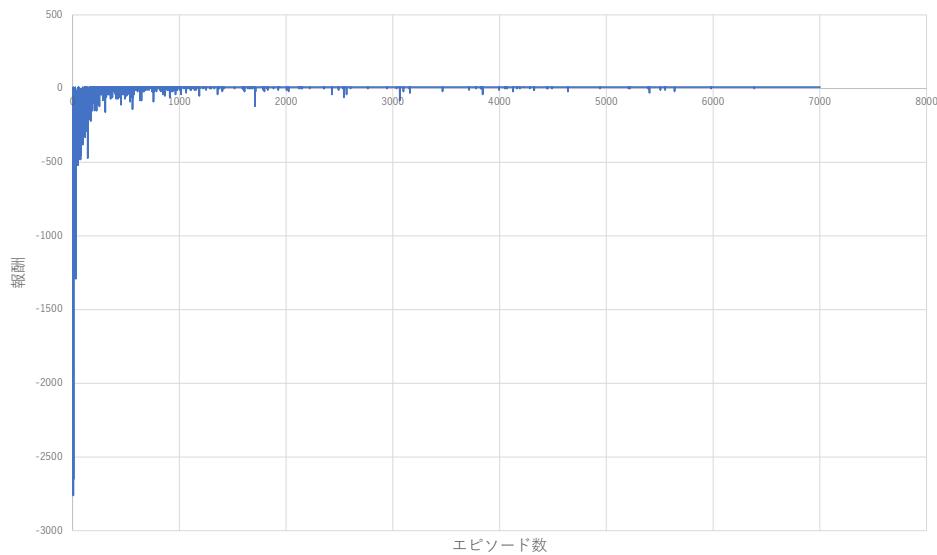


Fig. 4.17: 獲得報酬の推移を表した学習曲線（6番）

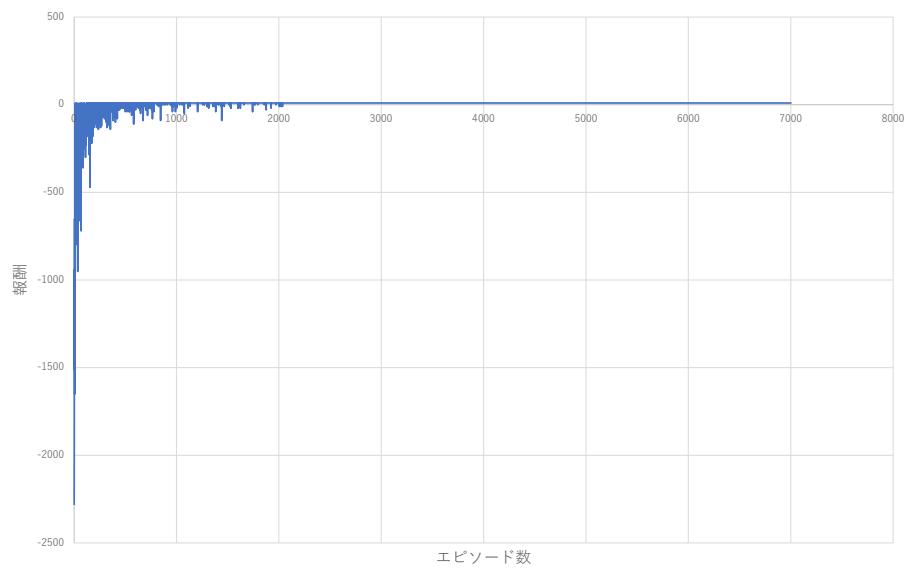


Fig. 4.18: 獲得報酬の推移を表した学習曲線（7番）

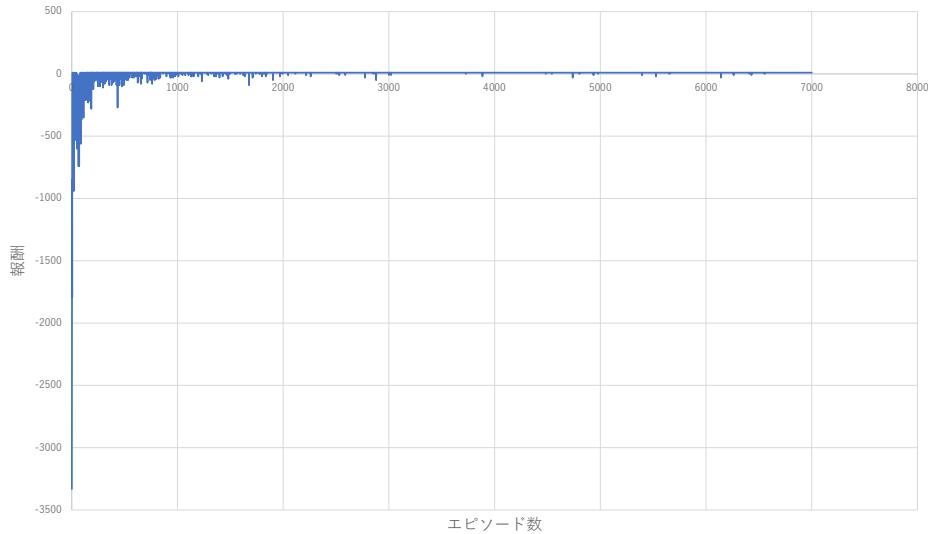


Fig. 4.19: 行動回数の推移を表した学習曲線（8 番）

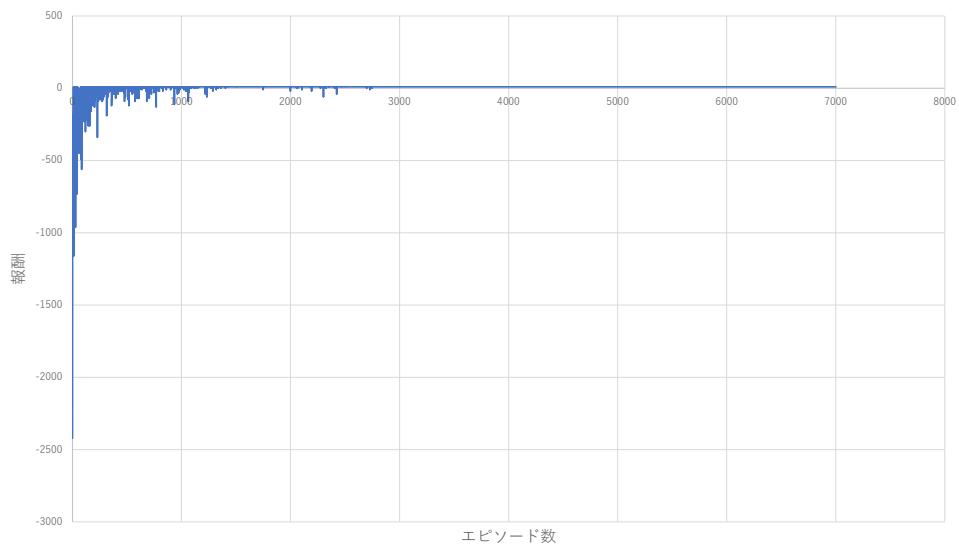


Fig. 4.20: 行動回数の推移を表した学習曲線（9 番）

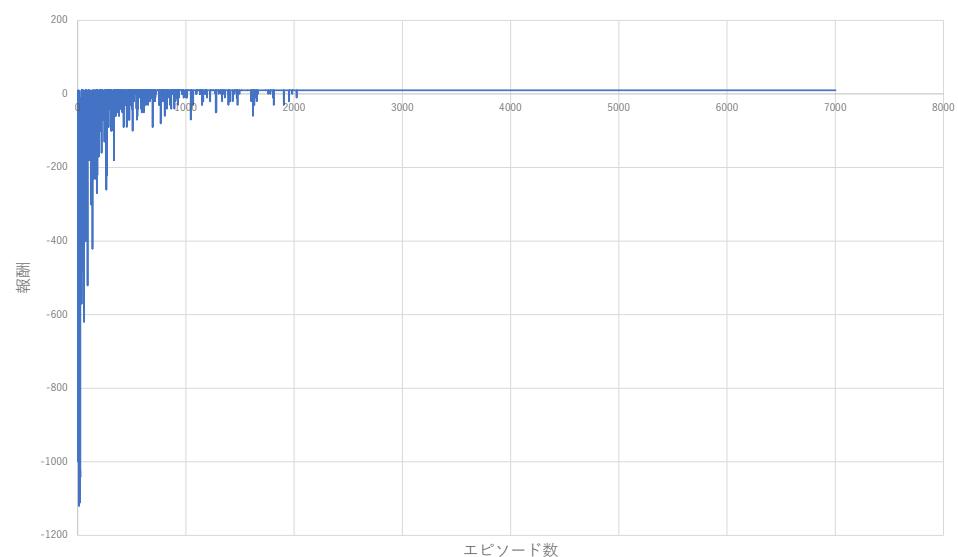


Fig. 4.21: 行動回数の推移を表した学習曲線（10番）

Fig.4.2～Fig.4.11 より、エピソード数が増えるにつれてステップ数が減少していることがわかる。このことからアルゴリズムの通り、探索しながら少ないステップ数で障害物を回避してゴールしていることを示している。Fig.4.12～Fig.4.21 より、エピソード数が増えるにつれて獲得報酬は正の値に収束傾向であることがわかる。このことから最適な行動戦略やポリシーを獲得し、報酬を最大化していることがわかる。以下に強化学習後のシミュレーションロボットを 1 エピソードだけ実行したので、10 パターンの障害物回避の移動軌跡を Fig.4.22～Fig.4.31 に示す。

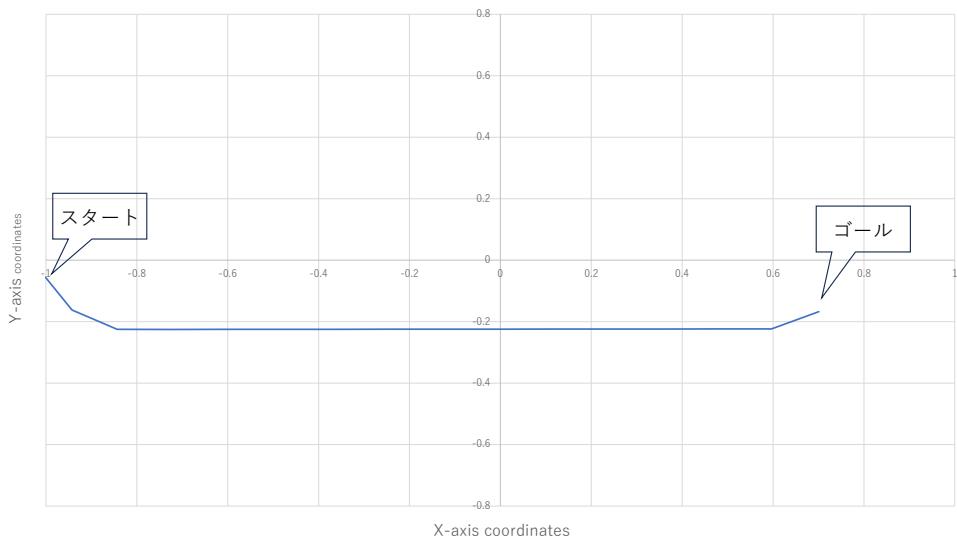


Fig. 4.22: 強化学習後の移動軌跡（1 番）

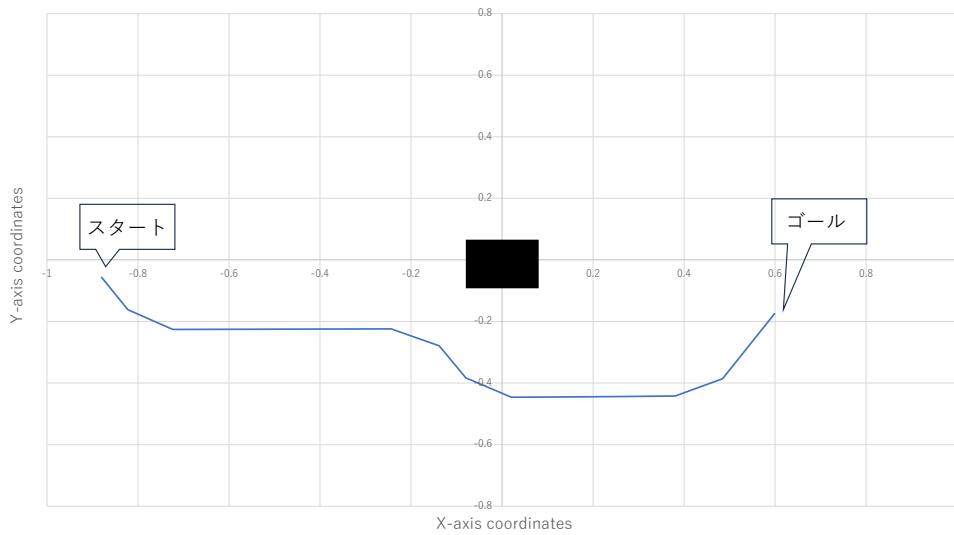


Fig. 4.23: 強化学習後の移動軌跡（2番）

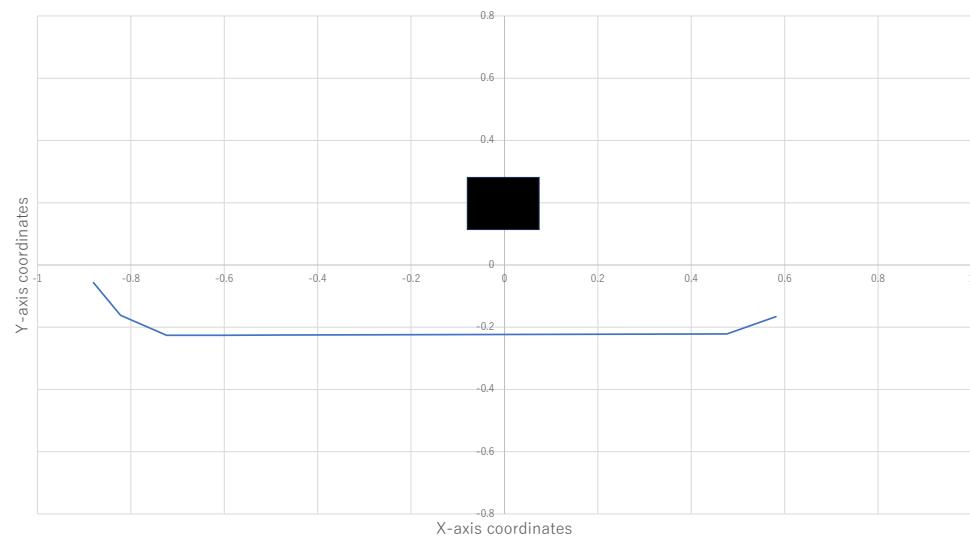


Fig. 4.24: 強化学習後の移動軌跡（3番）

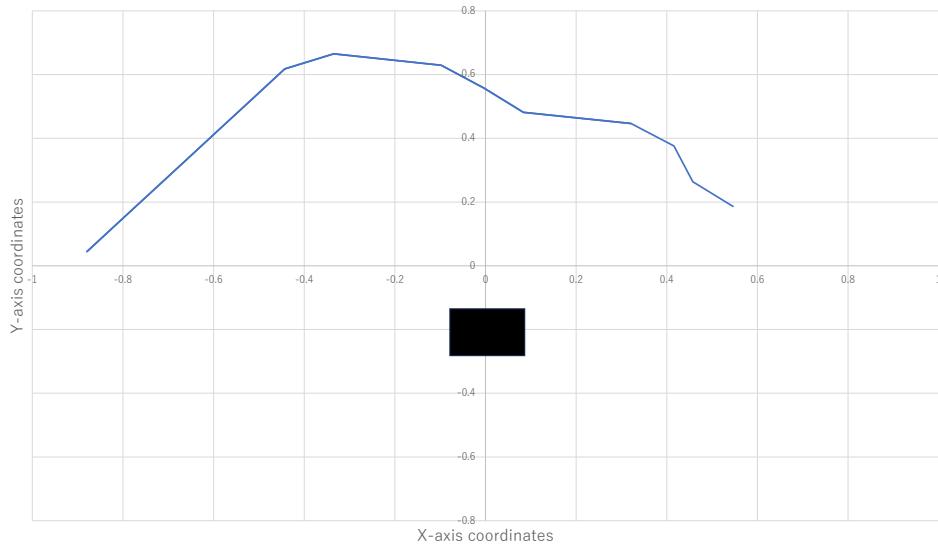


Fig. 4.25: 強化学習後の移動軌跡（4 番）

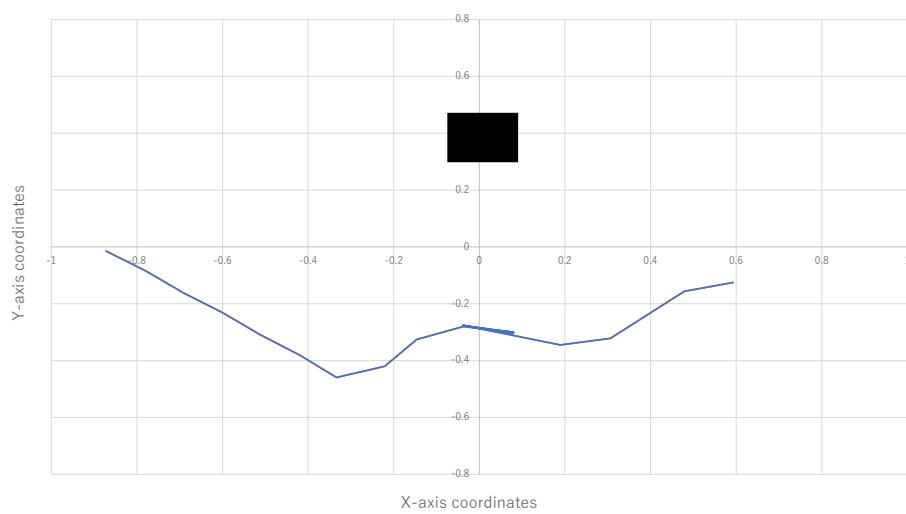


Fig. 4.26: 強化学習後の移動軌跡（5 番）

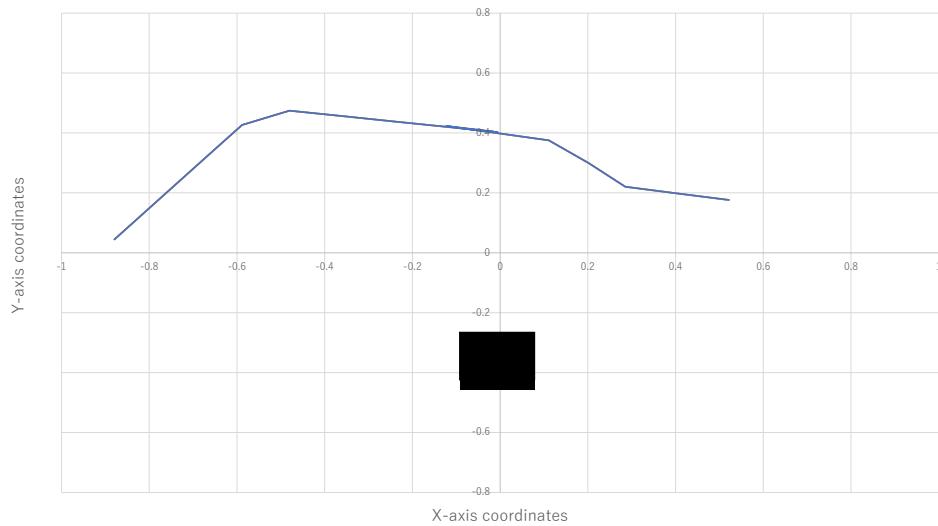


Fig. 4.27: 強化学習後の移動軌跡（6番）

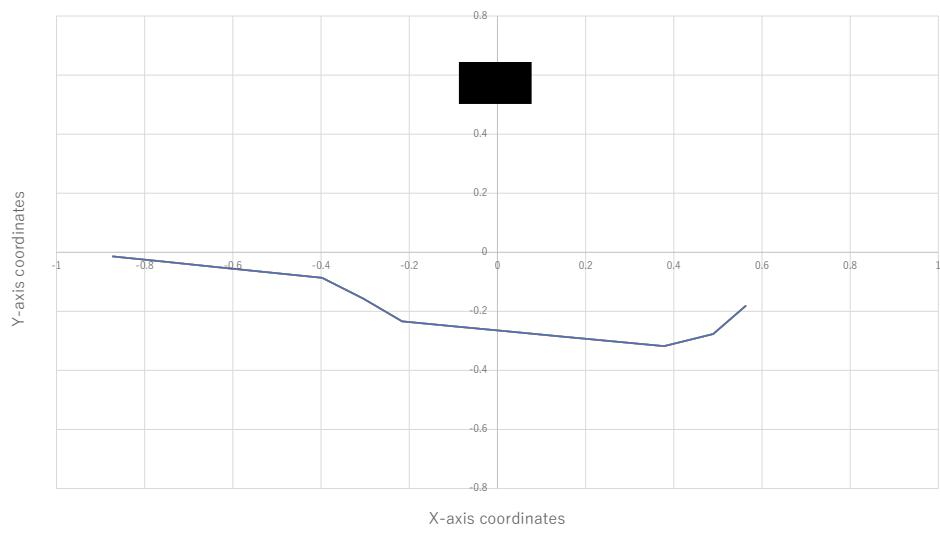


Fig. 4.28: 強化学習後の移動軌跡（7番）

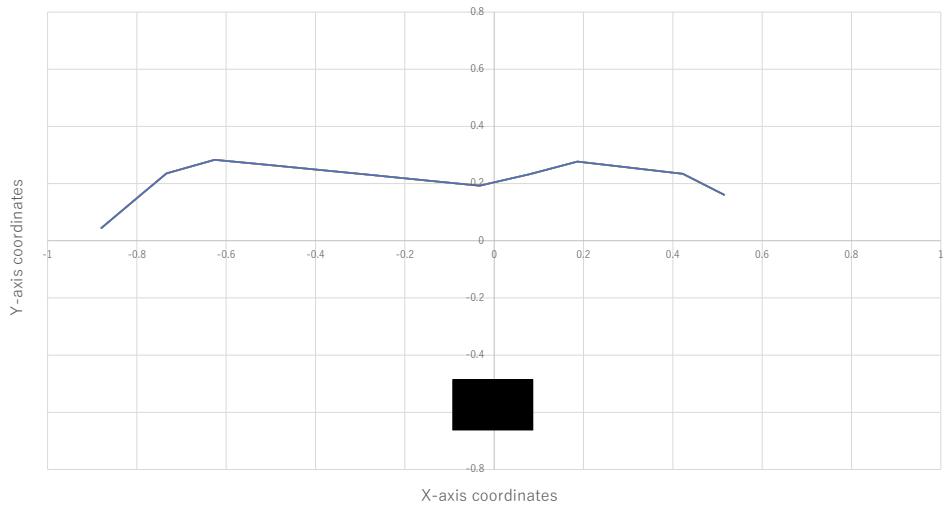


Fig. 4.29: 強化学習後の移動軌跡（8 番）

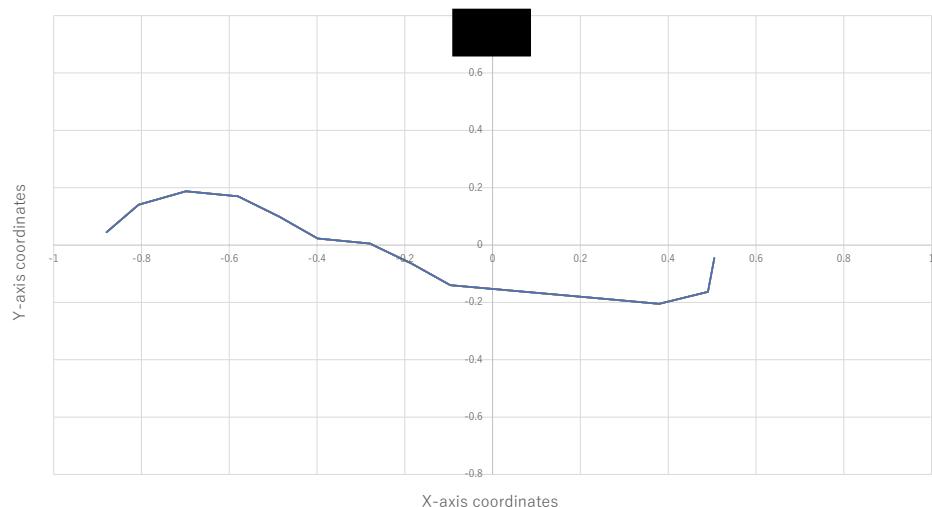


Fig. 4.30: 強化学習後の移動軌跡（9 番）

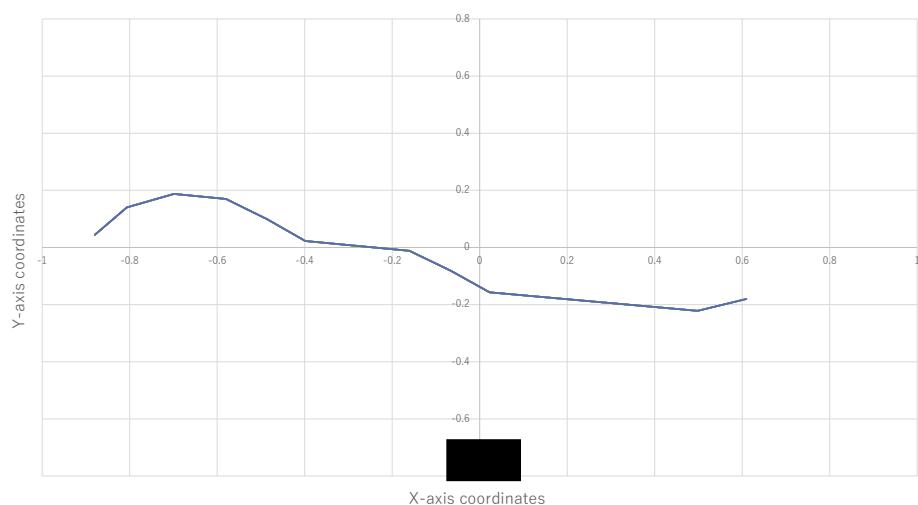


Fig. 4.31: 強化学習後の移動軌跡（10 番）

Fig.4.22～Fig.4.31 より、それぞれの配置パターンで障害物を回避するような移動軌跡が示せた。よって障害物を回避する行動策を学習することができた。また 10 パターンの障害物回避の強化学習の結果から、Fig.4.1 の配置を参考に、Fig.4.32 に示すような知識番号と知識間の重みが書かれた SAP-net の知識（方策）ネットワークを作成した。例えば障害物が前方にない場合は、1 番が選択されて直進をするような知識が選択される。障害物が 5 番の位置にある場合、知識番号 5 番が選択されて障害物を右に回避するような行動をとる。つまり、Fig.4.26 のような障害物を右に回避するような行動をとる。障害物が 4 番の位置にある場合、4 番が選択されて左を回避するような行動をとる。つまり Fig.4.25 のような障害物を左に回避するような行動をとる。

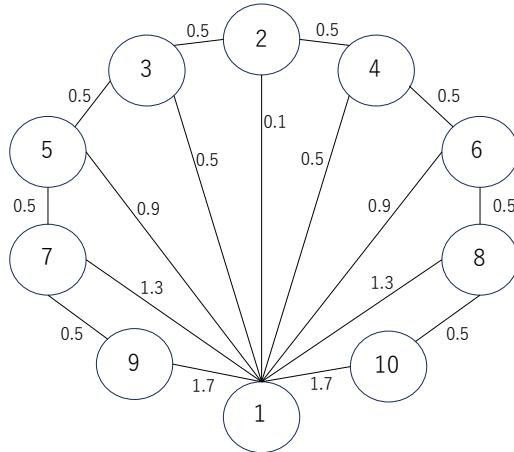


Fig. 4.32: SAP-net の知識（方策）ネットワークのイメージ図

4.3 SAP-net を実装したシミュレーション実験

4.3.1 目的と実験条件

前節にて述べた障害物回避の強化学習実験においては、Webots を用いたシミュレーション環境で Q 学習を基にした基本的な障害物回避の学習プロセスを実施し、ロボットが異なる障害物配置パターン下での効率的な回避行動を学習することができることを示した。これらの実験を通じて得られた知識は、ロボットが複雑な環境下での障害物回避能力を向上させるための基礎となる。本節では前節で得られた 10 個の知識（方策）を SAP-net の知識ネットワークの基盤として、さらに進んだ学習アルゴリズムの適用を行う。具体的には、SAP-net を通じて、ロボットが障害物までの角度と距離、ロボットの初期座標を考慮した上で、これまでに学習した知識（方策）と現在の環境との類似度を計算し、最適な行動選択を行うシミュレーションを実装する。障害物を中心に配置した環境でシミュレーションを行う。

4.3.2 実験結果

Fig.4.33 に行動回数の推移を表した学習曲線を、Fig.4.34 に獲得報酬の推移を表した学習曲線を示す。

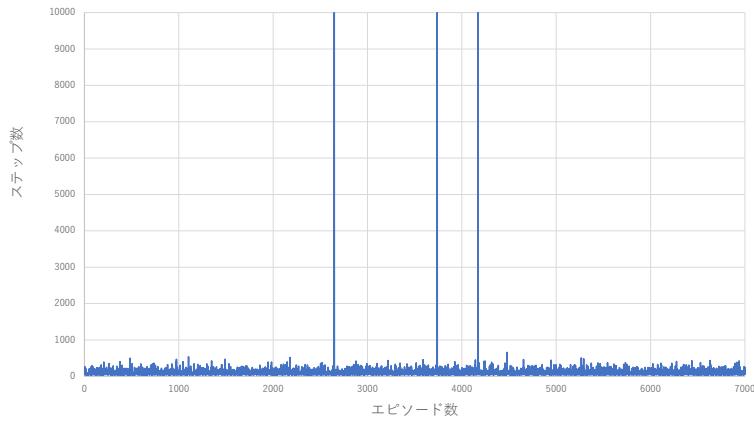


Fig. 4.33: 行動回数の推移を表した学習曲線

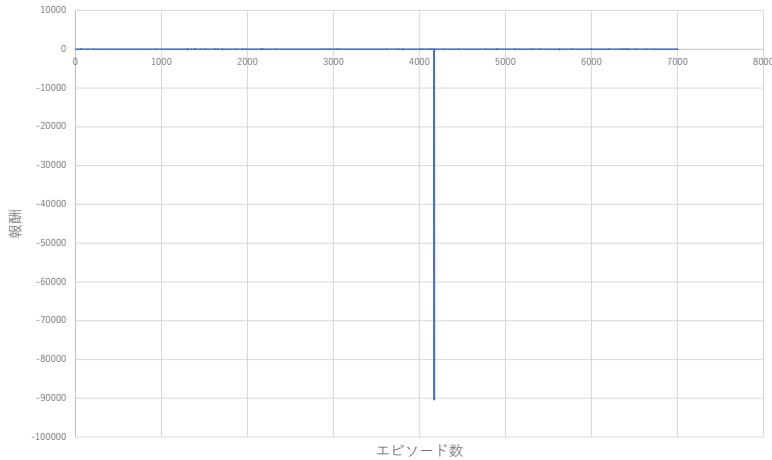


Fig. 4.34: 獲得報酬の推移を表した学習曲線

Fig.4.33 より、知識（方策）を保持した状態なので、強化学習実験とは違ってより少ない行動でゴールしていることがわかる。Fig.4.34 より、Webots でのバグが生じたものの、ほぼ正の値に収束していることがわかる。よって知識選択型転移強化学習を用いることで障害物回避の有用性を示した。また Fig.4.35 のような、障害物を 2 つ配置した場合の行動も検証したので、1 エピソードだけ実行して、その時の移動軌跡を Fig.4.36 に示す。

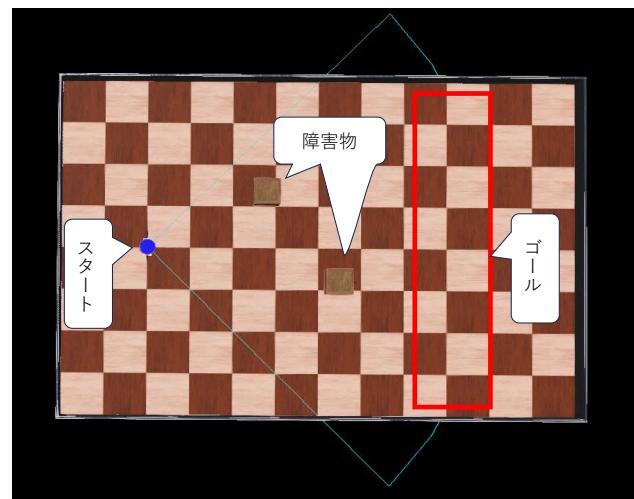


Fig. 4.35: 障害物を 2 つ配置した場合のシミュレーション環境

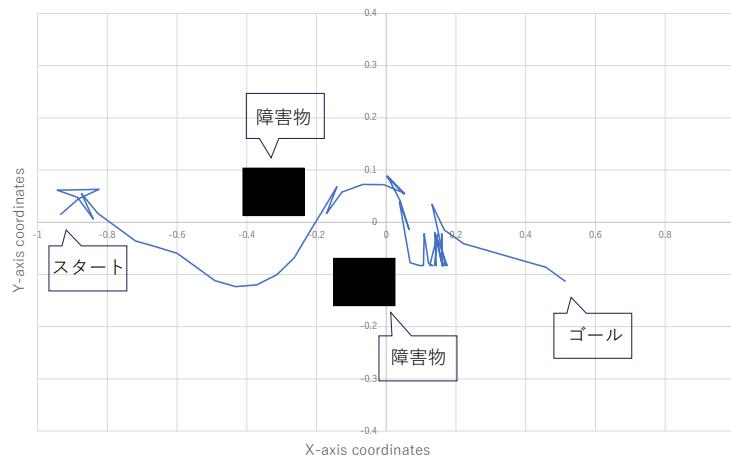


Fig. 4.36: 障害物を 2 つ配置した場合の移動軌跡

2 つの障害物を LIDAR が認識しながら障害物を回避していることが示せた。よって知識選択型転移強化学習を用いることで複数の障害物も回避することの有用性を示した。

4.4 実機実装と障害物回避性能の検証

4.4.1 目的と実験条件

本節では前節で得られた 10 個の知識（方策）と SAP-net を実機である株式会社 RT 製ラズベリーパイマウスに転移させて、前節と同じような環境で障害物回避の実現を目的とする。具体的には Webots シミュレーションで得られた 10 個の知識（方策）と SAP-net を、ラズベリーパイマウスという実機に転移させることで、実際の物理環境で障害物回避戦略の有効性を検証する。この実機実装の過程には、シミュレーションから実機へのパラメータ調整を含む、ラズベリーパイマウスに SAP-net アルゴリズムを組み込み、実環境での障害物回避の実現を目指す。実機実験では、シミュレーション環境と同等の条件下で物理的に障害物の配置を再現し、ラズベリーパイマウスが効率的に障害物を回避しつつ目標地点に到達できるかを評価する。使用する実機は Fig.4.37 に示すように、Slamtec 製 RPLIDAR を接続したラズベリーパイマウスを使用する。また、実験エリアを Fig.4.38 示す。



Fig. 4.37: ラズベリーパイマウス

4.4.2 実験結果

Fig.4.39 に障害物を右に回避した時の比較動合成画像を、Fig.4.40 に障害物を左に回避した時の比較動合成画像を示す。

Fig.4.39 および Fig.4.40 より、知識を選択しながら障害物を回避することに成功した。ま

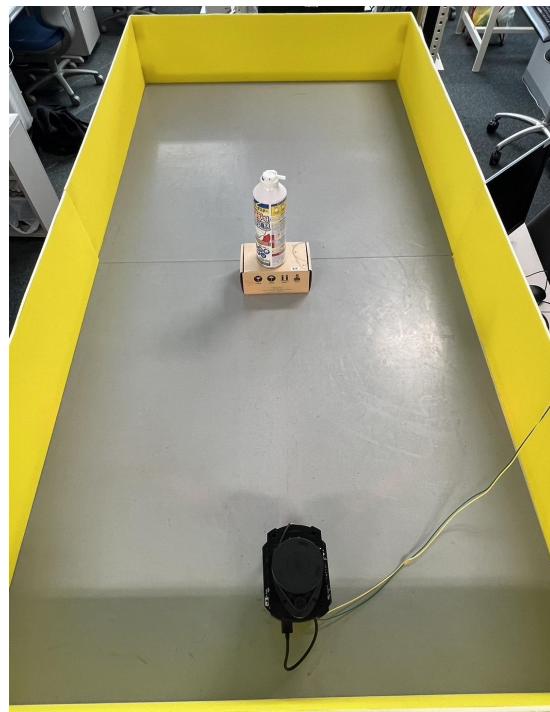


Fig. 4.38: 実験環境

た, Fig.4.32 から, ステップ数においての知識番号選択のグラフを Fig.4.41 と Fig.4.42 に示す.

Fig.4.41 より障害物を右に回避する際, Fig.4.32 の 3 番や 9 番が選択されていることがわかる. 3 番や 9 番は障害物を右に回避する際に選択されやすくなる番号である. また Fig.4.42 より障害物を左に回避する際, 6 番や 8 番が選択されていることがわかる. 6 番や 8 番は障害を左に回避する際に選択されやすくなる番号である. このことから環境情報を入力として知識を選択しながら障害物を回避していることが示せた. また, 初期位置でのセットアップで方向の誤差により, 進行方向が左右に少しずれて, 選択する知識が変更になり, 同時に障害物を回避する方向が変わったことが示せた.

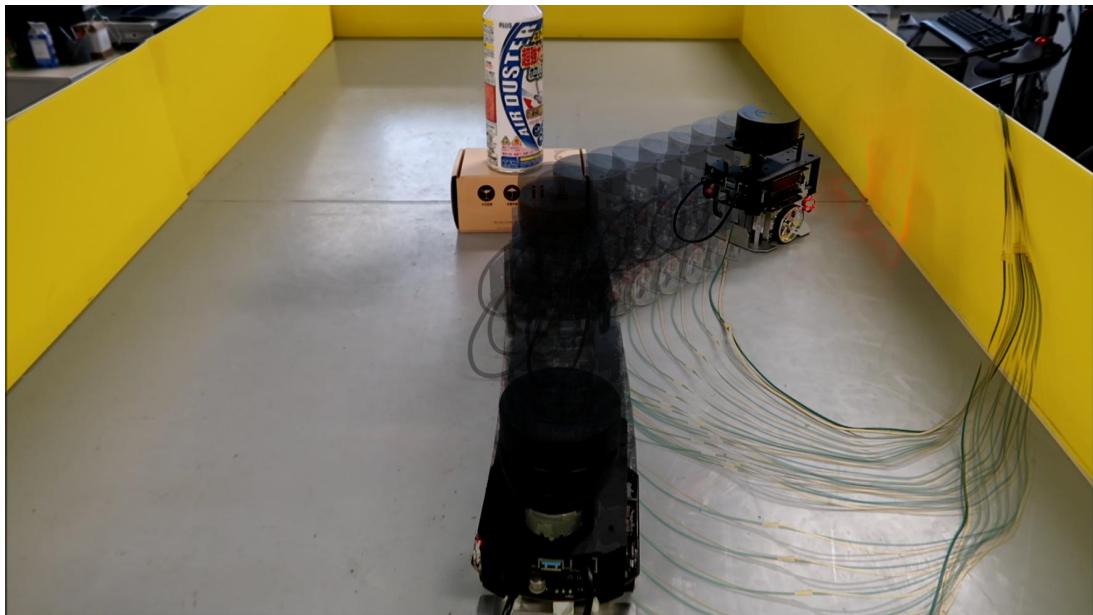


Fig. 4.39: 障害物を右に回避した時の比較動合成画像

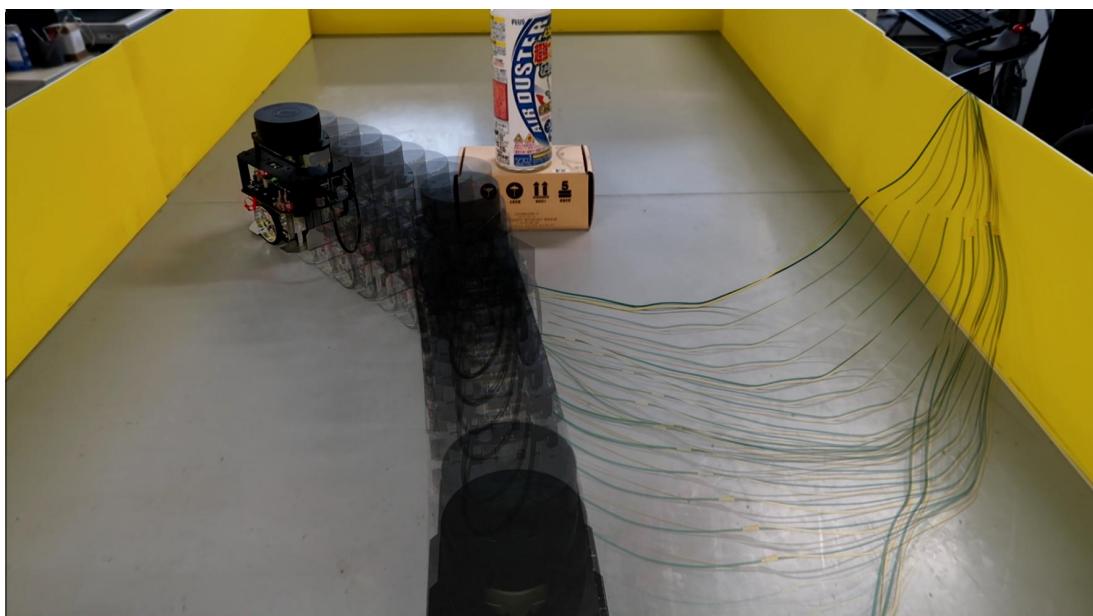


Fig. 4.40: 障害物を右に回避した時の比較動合成画像

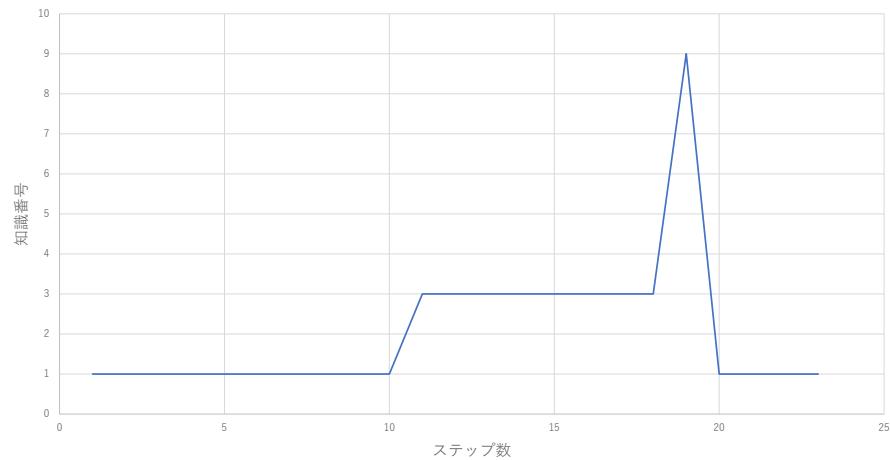


Fig. 4.41: 障害物を右に回避した時の知識選択の推移

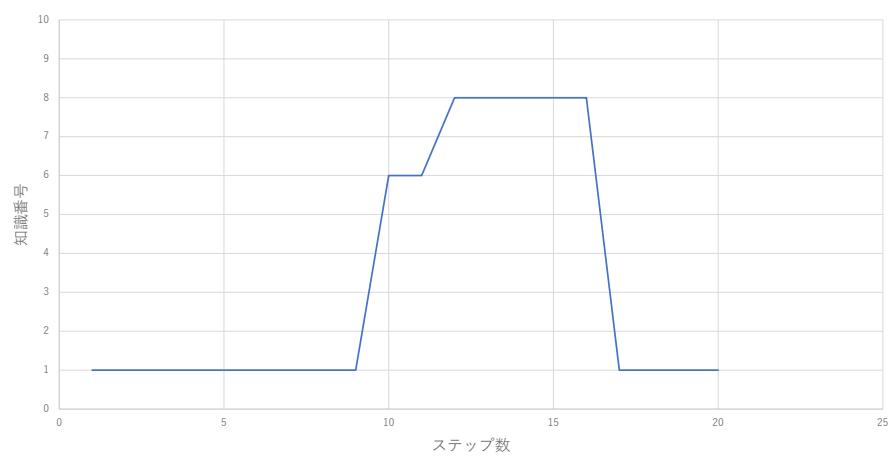


Fig. 4.42: 障害物を左に回避した時の知識選択の推移

4.5 おわりに

本章では、提案手法に基づいた実験の詳細について述べた。4.2 節では、強化学習実験の目的、条件、および結果について述べた。4.3 節では、SAP-net を実装したシミュレーション実験の目的、条件、および結果について述べた。4.4 節では、強化学習で獲得した知識を実機に転移させ、また SAP-net を実装させることで、実環境で知識を選択しながら障害物を回避することの有用性について述べた。本実験環境において、強化学習を用いた障害物回避の有用性、SAP-net を用いて複数の障害物の環境下を含めた障害物回避の有用性、実機実装を用いての障害物回避が可能であることを示した。

第5章

結論

Contents

5.1	結論	54
5.2	今後の展望	55

5.1 結論

近年自動車業界は AI や IoT などの先端技術によって大きな変革を迎えており、その中心には自動運転技術がある。特に、2023 年には日本で自動運転レベル 4 が解禁され、2025 年までに完全自動運転の実現が目指されている。自動運転の実用化により、安全性向上や交通事故の低減などの効果が期待されている。ただし、自動運転技術の発展と並行して、自動運転システムが搭載された車両による事故も発生している。事故を防ぐために LIDAR などの距離センサを用いて障害物を回避するような経路を生成するなどの研究はされているが、ルールベースによるプログラムのため、あらゆる環境や条件での障害物回避では機能しない。そこで本研究では、障害物を回避する強化学習を複数パターン実行して、知識数を増やしてあらゆる場面での障害物回避を提案した。第 1 章では、本研究の背景として自動運転技術の現状と事故事例、自動運転に関する関連研究をあげ、課題や問題点を提起、本研究の目的を「知識選択型転移強化学習を用いた自律型移動ロボットにおける障害物回避」とした。第 2 章では、本研究で必要となる、強化学習、転移学習、行動選択、知識選択の学習アルゴリズムの基礎知識について述べた。第 3 章では、本研究の目的を達成するための提案手法について述べた。第 4 章では、第 3 章で述べた提案手法の有用性を示すために、物理演算シミュレーションを用いた実験、実環境での障害物回避実験を行った。実験結果から Q 学習による強化学習のシミュレーション実験より、障害物回避の行動策を効率的に学習できたこと、SAP-net を実装したシミュレーション実験より、SAP-net で知識を保持しておくことで、障害物回避に対する最適な行動が実現できたこと、実機実装の実験では知識を選択しながら障害物回避を実現することができた。これから本研究において、障害物回避が強化学習で有用であること、多くの知識数での障害物回避が有用であることが示せた。以上のことから本研究における提案手法の有用性が示せた。

5.2 今後の展望

今後の展望としては以下の 2 つがあげられる.

1. 様々な環境での障害物の回避.
2. その他のロボットの適応.

本研究の目的は「知識選択型転移強化学習を用いた自律型移動ロボットにおける障害物回避」としている。異なる環境条件下での障害物回避能力を高めることが不可欠である。例えば、地形の多様性や様々な障害物の形状や配置など、多種多様なシミュレーション環境の条件が求められる。次に、四輪ロボットや大型のロボットといった、私たちが日常的に利用可能な様々な形態のロボットにて発揮するべきシステムであると考える。これにより、自動運転車だけではなく、配送料ロボット、産業用ロボットといった多岐にわたる用途での実用性が高まる。また工業施設、災害現場といった、予測不可能な障害物が存在しやすい環境下での運用を考慮することで、ロボットの自律性と安全性を大きく向上させることが可能となる。

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電機大学 工学部 情報通信工学科 河野 仁 准教授には、強化学習や Q 学習の構成、プログラムの構成等、専門性の高い分野を終始熱心なご指導と適切な助言を頂きました。この経験は極めて意義深く、将来において大いに役立つと確信しています。心から感謝申し上げます。

秘書の下山 芽衣さんには、研究活動において必要な事務作業が進められたおかげで、より効果的に研究を進めることができました。感謝申し上げます。

同輩である菊間 智陽君、江崎 皓平君、大須賀 日向君、久米 丈司君、山崎 拓真君とは違うグループではありますが、研究会の発表時に質問されたり、私の研究に対して鋭い指摘をされることで、私では気付かなかつたことにも気付かされて、私の研究をより充実させるきっかけとなりました。感謝いたします。

同じく同輩である小林 瑞樹君、佐野 康太郎君、末次 恭也君、高矢 空君、鳥谷部 悠希君とは、研究室で同じグループに所属しており、グループミーティングの際には私の研究に対して色々なアイデアを出してもらい、時には私の研究を補助してくれました。感謝いたします。

特に末次 恭也君、高矢 空君、鳥谷部 悠希君には、プログラムに関する助言や理解に困る点があれば、いつも親身になって教えてくれました。また日常生活においてもお互いにサポートし合い、精神的な支えとなる存在でした。彼らとの交流が私の人生に多くの豊かさをもたらし、心から感謝いたします。

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。

令和 6 年 2 月 須賀哉斗

参考文献

<和文文献>

[中川 2005]

中川 真仁, 安信 誠二: “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,” 第 21 回ファジィ システム シンポジウム, pp. 838–841, 2005.

[敷島 2021]

敷島 悠也, 田崎 豪: “単眼カメラと三次元地図を用いた動的障害物の検出と三次元復元,” 計測自動制御学会論文集, vol. 57, no. 1, pp. 37–46, 2021.

[金原 2022]

金原 翔, 米谷 昭彦: “自律走行車の軌道生成における不確定な動的障害物への対処方法,” 2022 年第 65 回自動制御連合講演会, pp. 462–467, 2022.

[河野 2022]

河野 仁, 坂本 裕都, 温 文, 藤井 浩光, 池 勇勲, 鈴木 剛: “知識選択型転移強化学習を用いたシニアカーの自律運転,” 2022 年電気学会電子・情報・システム部門大会, no. 1151, pp. 714–718, 2022.

[勞 2018]

勞世?, 陳謙: “自動運転システムにおける AI 技術,” 計測と制御, vol. 57, no. 7, pp. 493–496, 2018.

[齋藤 2014]

齋藤 碧, 小林一郎: “強化学習における効率的な転移学習適用に関する一考察,” 2014 年度第 28 回人工知能学生全国大会, no. 3, pp. 1–3, 2014.

[坂田 2022]

坂田 悠馬, 長尾 確: “強化学習における仮想環境と実環境における自動走行車いすの障害物回避,” 第 84 回全国大会講演論文集, pp. 81–82, 2022.

[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.

[米国家運輸安全委員会 2019]

米国家運輸安全委員会: “事故を起こした Uber Technologies の自動運転車,” <https://japan.cnet.com/article/35145765/>, 2019, 閲覧日 2023.12.20.

[ROHM 2020]

ROHM SEMICONDUCTOR: “ADAS. 自動運転, 安全を守る自動車のテクノロジーを解説,” <https://www.rohm.co.jp/blog/-/blog/id/8030502>, 2020, 閲覧日 2023.12.15.

[国土交通省 2015]

国土交通省: “自動運転を巡る動き,” <https://www.mlit.go.jp/common/001155023.pdf>, 2015, 閲覧日 2024.1.15.

[厚生労働省 2023]

厚生労働省: “厚生労働省「人口動態統計（確定数）/2022」,” https://www.mhlw.go.jp/toukei/saikin/hw/jinkou/kakutei22/dl/15_all.pdf, 2023, 閲覧日 2024.1.15.

<英文文献>

[Taylor 2009]

Matthew E.Taylor and Peter Stone: “Transfer Learning for Reinforcement Learning Domains:ASurvey,” *Journal of Machine Learning Research 10*, pp. 1633–1685, 2009.

[Kono 2019]

H. Kono, R. Katayama, Y. Takakuwa, W. Wen, and T. Suzuki: “Activation and Spreading Sequence for Spreading Activation Policy Selection Method in Transfer Reinforcement Learning,” *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 12, pp. 7–16, 2019.

[Webots 1998]

Cyberbotics, Webots OPEN SOURCE ROBOT SIMULATOR, <https://cyberbotics.com>, 1998.

研究業績

査読有り国内会議

1. 高矢 空, 河野 仁, 須賀 哉斗, 鳥谷部 悠希, 池 勇勲, 藤井 浩光, 鈴木 剛: “知識選択型転移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,” 2023 年電気学会電子・情報・システム部門大会, 2023, pp.933–936, 北海道.