目次

第 1 章　序論

1.1 ロボットが使用される環境とその変化

実用初期のロボットは，人間によって事前に設定された動作を繰り返すものが主流であり，工場での組み立て作業などに用いられていた．そのため，ロボットが使用される環境は限定された場所であり，ロボットにとって最適化されていた．例えば工場の組み立てラインで作業するロボットは，工場のラインという限定された環境で，人間によって設定された組み立て作業を繰り返し行うのみである．現在でもこのような産業用のロ移動機能，センシング機能などが向上した．この結果ロボットが使用される環境は，様々な場所へ広がることとなった．

現在，ロボットが活躍する場は自然環境，家庭環境，オフィスなど以前より人間の生活に近い場所へと広がってきている（図 1.1）．人間の生活環境は，実用初期のことロボットが使用されていた環境に比べて複雑であり，かつ時間によって環境変化が発生する．人が住む家を例に挙げると，配置されている家具や雑貨の位置は日々変化する． 一度ある場所にあったものが，そのままの場所にあるとは限らない．人間が生活する場所ではある日テーブルの上にあった雑誌が，次の日には床に置いてある可能性がある．

このように，ロボットが使用される環境は整備された環境から，複雑で多様な環境へと変化してきている．

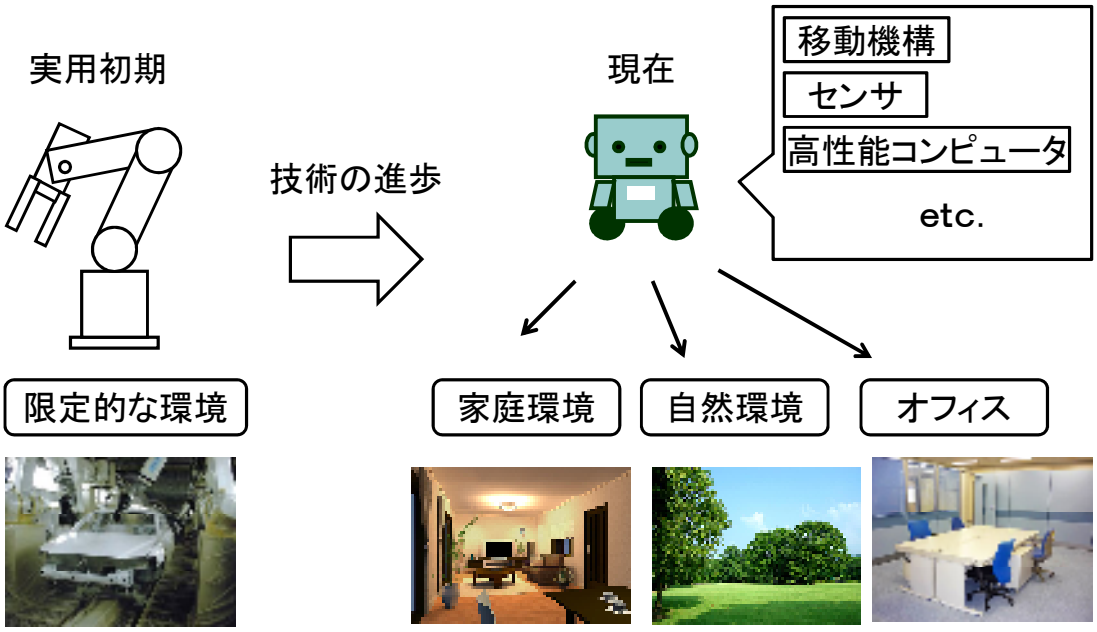


図 1.1　ロボットが使用される環境の変化

1.2 学習機能によるロボットの環境への適応

ロボットの実用初期のように限定された環境においてはロボットが直面する環境やロボットの状況を人間が事前に予測することが可能であった．従って，環境に対するロボットの行動を設定することが可能であった．しかし，家庭環境など変化が多い複雑な環境下においては，直面し得る環境全てを人間が予測することはできない．そのためロボットが直面する状況に対して行うべき行動を人間が事前に設定することは難しい．そこで，複雑な環境下ではロボットが自律的に周りの環境に応じた行動を取ることが求められる．ロボットが自律的に周りの環境に応じた行動を取るための方法の一つとしてロボット自身に学習機能を持たせる方法がある[1][2].

ロボットの学習機能とは，人間が以前の経験をもとに未知の環境に対しても適応可能であるのと同様に，ロボットにも以前の経験から周りの環境に合わせた行動を行えるような知能を持たせることである．人間が事前に行動を選択するための規則、ルール、判断基準などを与えることで，ロボットは未知の環境であってもその環境に応じた行動を取ることが可能となる.

このようにロボットは，学習機能を有することで変化する環境に対してある程度適応することが可能となった.

1.3 学習機能を持つロボットにおける身体状態の考慮

学習機能を持つロボットに限らず，ロボットは行動するために身体が必要となる．ロボットの身体はアームや駆動輪，モータ，回路，バッテリーなどから構成されており，身体を構成する要素に問題があった場合には活動することができない．例えば，移動ロボットは駆動輪が故障した場合には移動することができなくなり，回路に異常が発生した場合やバッテリーが不足した場合には機能停止となってしまう.

学習機能を持つロボットは，人間に与えられた仕事の遂行のみを目標として行動する．よって，ロボットの身体に問題があった場合でもロボットは人間に与えられた仕事の遂行を優先することとなる．そこでロボットは身体の状態を考慮した行動を行うことが必要となる．先に述べたように，学習機能を持つロボットは人間によって事前に与えられた指標に基づいて仕事の遂行を行う．よって，人間がロボットの身体の状態を考慮しない指標を与えた場合には，ロボットが故障や機能停止を引き起こす可能性がある．例えばバッテリーを電源とする自律ロボットの場合，バッテリー残量が少ないときに人間から与えられたた仕事を実行しようとしても電力不足で仕事の遂行が行なえず機能停止に陥るといったことが起こり得る（図 1.2）.

このような問題を解消するためには，人間が事前に身体状態を考慮してロボットの行動を設定する必要がある．ロボットの身体はホイール，アーム，フレーム，回路，セン

サ，バッテリーなど多くの要素によって構成されており，ロボットによってその要素も異なる．よって，ロボットの全ての要素を考慮することは難しい．

学習機能を持つロボットに限らず，ロボットは行動，計算，センシングなどに電力を必要とし，電力なしでは活動することができない．そこで本研究では，ロボットが電力を獲得するために搭載するバッテリーの状態に着目する．

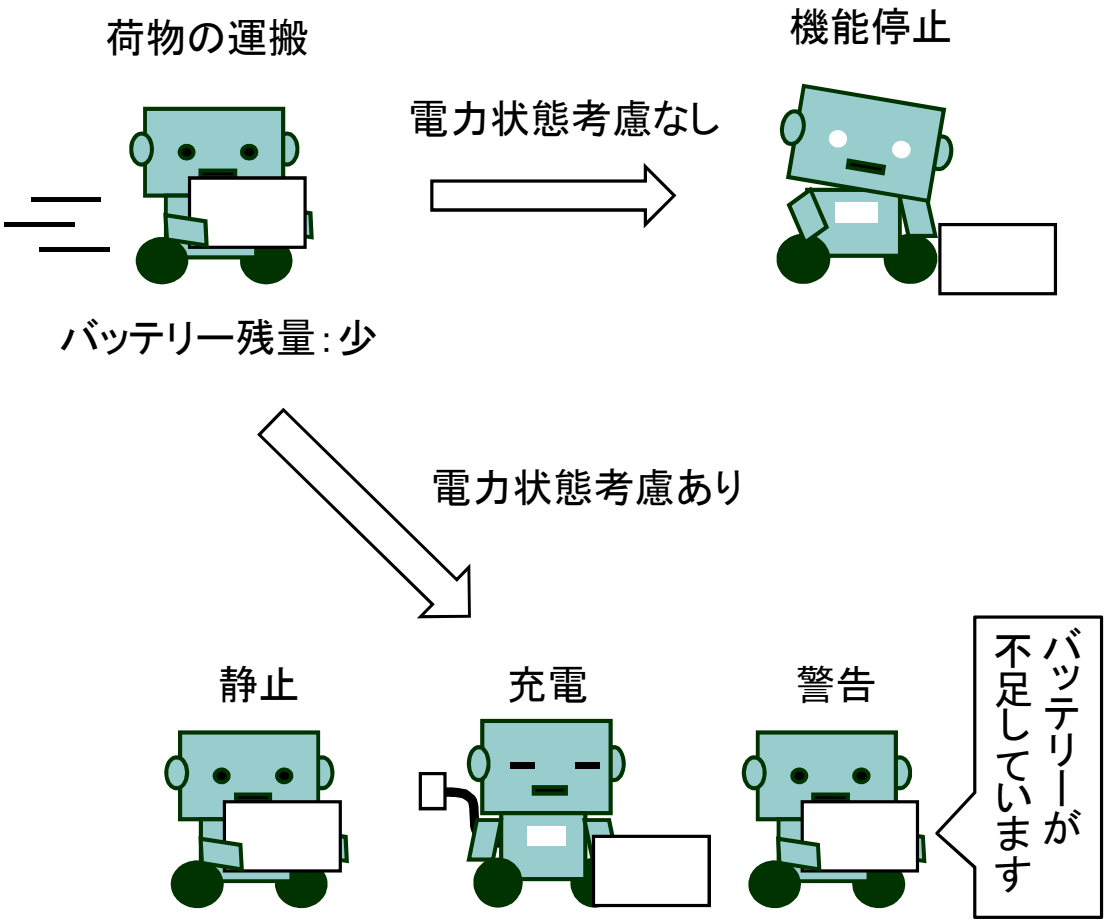


図 1.2　電力状態の考慮の有無

1.4　従来研究と問題点

自律的に行動を選択するロボットにおいてロボットの電力状態を考慮した従来研究がいくつかある[3]-[6]．“律移動ロボットの自己保存-自律移動，6 軸力覚センサを用いたコンセント挿入動作の実現”という研究では，自律移動ロボットに充電機能を搭載している．この研究では，自律移動を行いスタートからゴールまでの経路を探索するロボットにおいて，バッテリー残量が一定の値を下回った場合に充電を行うことで機能停止を回避している．扱われている自律ロボットは家庭コンセントに接続可能なプラグを搭載しており，自律行動中にコンセントの位置を発見して自律的な充電を行うことを目標としている．また，“エネルギー自給型屋外環境ロボット”という研究においては，自律型の芝刈り機に太陽電池を搭載しロボットのバッテリーが無くなることで機能停止することを回避している．この研究では，ロボットに搭載されたバッテリーの充電状況が「不良」と判定された場合に太陽電池によって充電を行う方法を取っている．

また，社会で普及している，ＡＳＩMOや家庭用掃除ロボットにおいても搭載されたバッテリー残量を考慮した設計が行われている．このような社会で普及しているロボットに設定されているバッテリー残量を考慮した一般的な方法としては，ロボットに搭載されたバッテリー残量が一定の値を下回った場合に充電器へ戻り充電を行うといった方法が取られている．

従来研究や社会に普及している自律ロボットにおいて一般的に用いられているバッテリー残量の考慮は，ロボットが機能停止を回避する行動を取るバッテリー残量の基準は，人間が事前に設定しているという点が共通している．この基準を事前に設定できるのは，ロボットが使用される環境を人間が事前に想定できるからである．例えば充電機能を搭載した掃除ロボットの場合，使用される環境は屋内であり，使用される環境の広さも十数メートル程度と想定することができる．環境が想定できれば，充電を行うための充電器までどの程度の距離なのかなどを考慮し機能停止を回避するためにはバッテリー残量が何%以下となるまでに充電を行えばよいか設定できる．

しかし，学習機能を持つロボットは，環境に対する行動を自律的に獲得できるという特性から，未知の環境で使用されることが多い．ロボットが使用される環境が，未知環境となった場合には人間が事前にロボットが直面する環境や状況を想定することはできなくなる．従って，従来研究のような一定の基準を設定することは難しいといった問題が起こる．この問題を解消するためには，ロボットが機能停止を回避するために行動を行うバッテリー残量の基準を使用される環境に合わせて動的に設定する必要がある．

本研究では，特に未知環境でのロボットの行動について注目する．未知環境では，ロボットが同じ行動を取った場合であっても，消費するバッテリーが異なる場合がある．例えば，一定の距離を走行するだけでも，走行する環境が平地か坂道かで消費するバッテリー量がことなる（図 1.3）．そのため，同じバッテリー残量でも，取る行動のバッテリー残量によって機能停止を回避する行動を取るかどうかが変わる．従って，ロボット

の行動によって消費するバッテリーの量に応じて，自律的に機能停止を回避するバッテリー残量の基準を変更する必要がある．

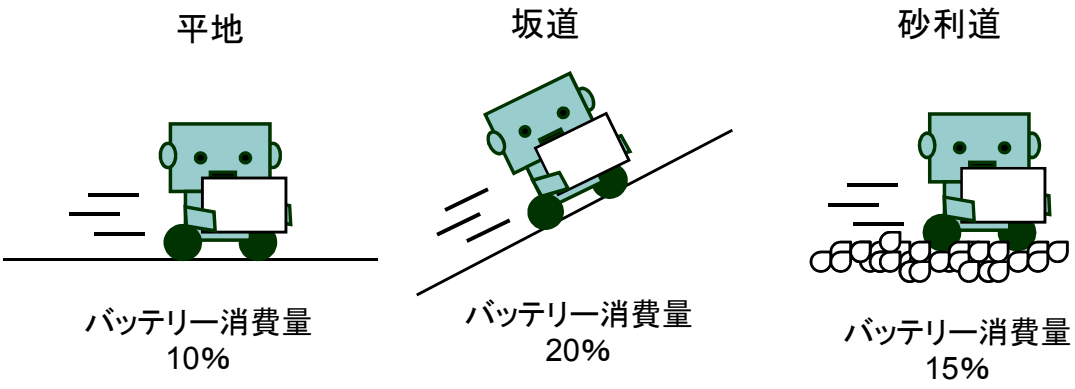


図 1.3　環境によるバッテリー消費量の違い

1.5 本研究の目的

本研究では，ロボットが活動するためには電力が必要であり，電力不足による機能停止を防ぐためにはロボットに機能停止を回避する行動を設定する必要があることに着目する．従来研究では，人間が事前に決定した基準で機能停止を回避する行動を設定していたが，未知環境においてはロボットが使用される環境や状況を事前に想定することはできず，その環境にあった基準を動的に設定する必要である．

そこで本研究では，ロボットの行動におけるバッテリー消費量に基づき，バッテリー残量に応じた行動選択を行う学習システムを提案する．提案する学習システムによって，バッテリー残量に対する行動を自律的に決定するロボットを実現する．ロボットの学習としては，最も一般的な学習手法である強化学習を用いて学習システムを設計する．

1.6 本論文の構成

第 1 章では，学習機能によって環境変化に適応するロボットについて説明し，ロボットは自身の身体状態を考慮して行動することが必要であることを述べた．そして，ロボットの身体状態のなかでもバッテリー残量について注目し，行動におけるバッテリー消費量に基づいて，バッテリー残量に応じた行動選択を行う学習システムを提案することを目的とした．

第 2 章では，本研究で対象とする強化学習について説明し，ロボットへ適用した場合にどのような流れで学習を行うのかを説明する．

第 3 章では，ロボットの自己保存について説明し，強化学習を適用したロボットとの関係を説明する．

第 4 章では，提案する学習システムについて説明する．初めにシステムの全体的な概要，構成を述べる．その後，提案する学習システムにおける学習の方法について詳しく述べる．

第 5 章では，本研究で提案する学習システムの検証実験について述べる．また，結果について考察する．

第 6 章では，検証実験の結果をもとに全体のまとめを述べ，今後の課題についても説明する．

第 2 章　強化学習

2.1 強化学習の概要

強化学習[7]とは，環境に対する試行錯誤を通じて適切な行動を獲得する機械学習の一種である．ロボットに強化学習を適用することで周りの環境に適応し，より良い行動を選択することが可能となる．

強化学習では，報酬と呼ばれるスカラ値を用いて学習を行う．ロボットが周囲の環境に対して行動することで環境から報酬を獲得することができる．ロボットと環境の関係図を図 2.1 に示す．ロボットは環境との試行錯誤を通して報酬が最も多く得られるような行動を選択する．ロボットが環境から与えられる報酬は人間が事前に設定する．従って，ロボットに行ってほしいタスクに応じて，報酬を設定する必要がある．

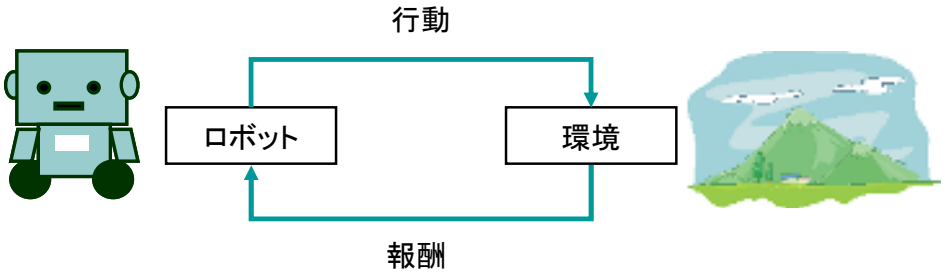


図 2.1　ロボットと環境の関係図

強化学習では，ロボットが行うタスクに対して何が最適な行動なのかを与える必要はない．人間がタスクに対してロボットに行ってほしい行動ほど高い報酬を与え，好ましくない行動に対して低い報酬を与えることで，ロボットは自動的に最善の行動を獲得することができる．従って，強化学習を適用したロボットは未知の環境を扱うことが可能であり，実ロボットに用いられることが多い．

2.2 強化学習における学習の流れと学習方法

前節では，強化学習の概要について述べた．本節では強化学習における学習の流れを説明し，具体的な学習方法について述べる．

強化学習で対象とするロボットにはセンサが搭載されており，センサを通して周囲の状態を認識することが可能である．また，ロボットは認識した状態に対して何らかの行動を取ることができる．このようなロボットに対する強化学習の概念図を図 2.2 に示す．

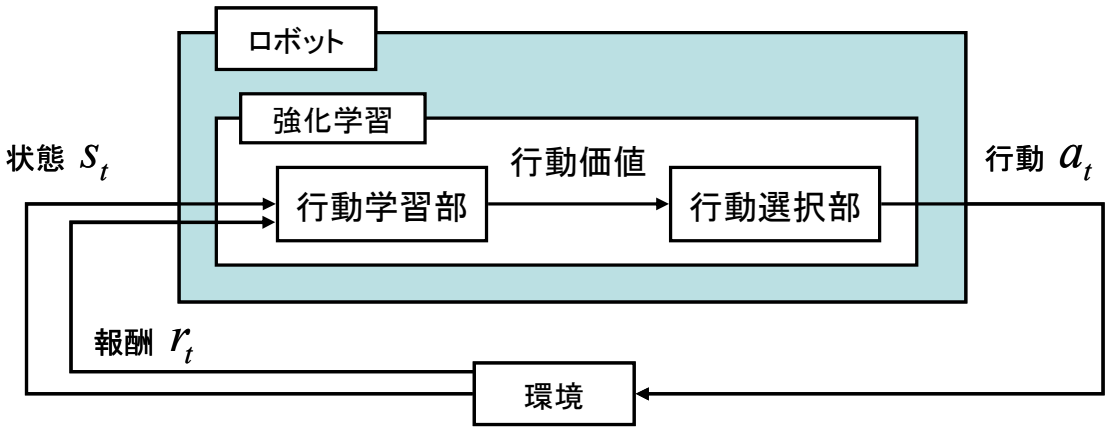


図 2.2　強化学習の概念図

時刻tにおいてロボットが認識した状態s_tをとする．ロボットは認識した状態s_tに対して，過去の学習結果から行動a_tを選択する．このとき選択した行動に対して環境から報酬r_tを獲得し，獲得した報酬r_tとロボットが認識した状態s_tは行動学習部に入力される．行動学習部では状態s_tと行動a_tの組み合わせに対する行動価値を更新する．この状態s_tと行動a_tの組み合わせを状態行動対と呼ぶ．行動価値とは，ロボットが取った行動によって獲得できる報酬の期待値である．報酬はロボットが取った行動に対して即時的な意味合いで良し悪しを示すものであるのに対し，行動価値は最終的な行動の望ましさを示している．行動価値の更新方法は，用いる行動学習法のよって異なるため詳しくは 2.3 節で説明する．行動学習部で算出された行動価値は行動選択部へ入力される．行動選択部では，行動学習部で更新された行動価値から次の行動を選択する．この行動の選択方法は，用いる行動選択手法によって異なるので詳しくは 2.4 節で説明する．ロボットはこのサイクルを繰り返すことでロボットは目的遂行のために最適な行動を学習することが可能となる．この流れをまとめた図を図 2.3 に示す．

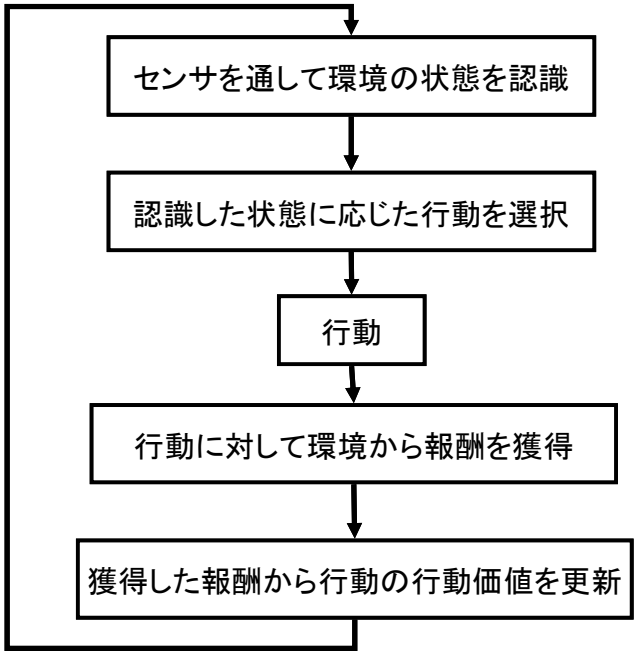


図 2.3　強化学習の流れ

2.3 行動学習手法

行動学習手法は 2.2 節で説明した行動学習部において, 環境から獲得した報酬から行動価値を評価する手法である．本研究では，行動学習手法として加重平均法を用いるので加重平均法について説明する.

加重平均法は，遠い過去に受け取った報酬を考慮するのか，近い時刻に受け取った報酬のみを考慮するのかを重み付けによって変更し行動価値を更新する方法である．ある時刻tにおいて状態s_tで行動aを取った場合の行動価値$Q_t(s.a)$の更新式を式(2.1)に示す. αはステップサイズパラメータと呼ばれる定数で，この値を変化させることによって与える重みが変化する．αの値を大きく設定した場合には，近い時刻に受け取った報酬の影響が大きくなり，αを小さく設定した場合には遠い過去に受けとった報酬も考慮した行動価値の更新となる.

$$Q_{t+1}(s,a) = Q_t(s,a) + \alpha[r_{t+1} - Q_t(s,a)] \tag{2.1}$$

$Q_t(s,a)$：時刻t，状態sで行動aを選択した場合の行動価値

r_{t+1}：新たに獲得した報酬

α：ステップサイズパラメータ（$0 < \alpha < 1$）

2.4 行動選択手法

行動選択手法は 2.2 節で説明した行動選択部において, 過去の行動価値から最適な行動を選択するための方法である．ここでは，代表的な手法として ε -greedy 法と softmax 法について説明する.

2.4.1 ε -greedy 法

ε -greedy 法は，過去の行動価値の中から ε の確率でランダムな行動を選択し，1− ε の確率で行動価値の最も高い行動を選択する行動選択の方法である（図 2.4）． ε の確率でランダムな行動を取ることで，現在最も高い行動価値となっている行動よりも更に良い行動があるかどうかを調べることができる.

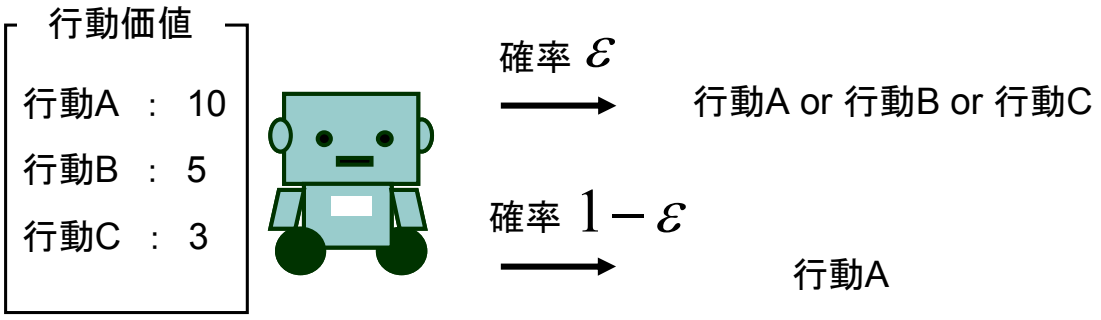


図 2.4　ε -greedy 法の概念図

2.4.2 softmax 法

softmax 法は，行動を選択する確率を設定する．この確率を行動確率という．softmax 法では，行動価値によって行動確率を変化させる方法である．つまり，行動価値の高い行動には最も高い行動確率が与えられ，その他の行動は，行動価値の高い順に行動確率が与えられる．時間 t において状態 s で行動 a を選択する確率 $\pi_t(s,a)$ は式（2.4.2）で与えられる． τ は温度と呼ばれる正定数である.

$$\pi_t(s,a) = \frac{e^{Q(s,a)/\tau}}{\sum_{b=1}^{n} e^{Q(s,b)/\tau}} \qquad (2.4.2)$$

$\pi_t(s,a)$ ： 時間 t，状態 s で行動 a を選択する確率

$Q_t(s,a)$ ： 時間 t，状態 s で行動 a を行った場合の行動価値

τ ：温度

温度τが高い場合には，全ての行動がほぼ同程度に選択される．また，温度τが低い場合には行動価値の高低による選択確率の差が大きくなる．

第 3 章 ロボットの自己保存と強化学習の関係

第 2 章では，ロボットの学習方法の一つである強化学習について説明した．本章では，強化学習を適応したロボットと自己保存機能との関係について説明する．第 1 節では，ロボットにおける自己保存の定義について述べる．次に，第 2 節で自己保存を考慮したロボットとタスク遂行のみを目標とするロボットの違いについて述べる．

3.1 ロボットにおける自己保存の定義

ロボットに自己保存機能を搭載するためには，まずはロボットの自己保存とはどのようなものなのかを定義する必要がある．自己保存とは本来，生き物が有するものであり，我々人間にも自身の生命を保存しようとする自己保存の本能がある．人間は身体が失われれば生きていくことができない．また，活動するために必要となるエネルギーが不足した場合にも同様である．このため人間はおなかが減ると食べ物を食べ，エネルギーを補給する．また，自分の身に危険が迫った場合には身の安全を確保しようとする．このような行動が人間における自己保存だと考えられる[8]-[10]．

この概念はロボットにも当てはめることができる．回路などの電子機器を用いないロボットを除いてほとんどのロボットは活動するために電力を必要とする．よって，ロボットは電力がなければ活動できず機能停止となってしまう．この問題を回避するためには，ロボットが自身の活動に必要な電力を維持しなければならない．そこで本研究では，活動に電力が必要なロボットを対象とし，ロボットが自身の活動に必要な電力を維持することを自己保存と定義する．

3.2 タスク遂行のみを目的とするロボットと自己保存を考慮するロボットの違い

ここでは，行動するために電力が必要であり，必要な電力をバッテリーによって確保しているロボットを対象として話を進める．強化学習によって人間に与えられたタスクを遂行することのみを目的とするロボットとこの目的のほかに自己保存も考慮して行動するロボットでは，搭載されたバッテリーの量に対して取る行動が異なる．強化学習によって人間に与えられたタスクを遂行するロボットは，タスクの遂行を目的とした報酬を与えられている．従って，人間がバッテリー残量を考慮した報酬関数を設定しない限りは，バッテリー残量に関係なくロボットはタスクの遂行を目指して行動を行う．荷物を目的地まで最短で運搬することを目的とする学習ロボットの場合，報酬は目的地への最短経路を通った場合に高く与えられる．従って，運搬中にバッテリー残量が少なくなった場合でも，ロボットは充電を行うといった行動を選択する余地はない．よって，荷

物の運搬中に機能停止を引き起こしてしまう．

一方，自己保存を考慮したロボットであればバッテリー残量が少なくなった場合に与えられたタスクの遂行とは別に自己保存を考慮することが可能である．従って，バッテリー残量が少ない場合には，人間に与えられたタスクの遂行より自己保存を優先し，充電を行ったり，人間に警告して電力も供給を要求したり，静止状態となり機能停止を防ぐなどの対策を取ることができる．先ほどの荷物を目的地まで最短で運搬することを目的とするタスクを例に取ると，自己保存を考慮したロボットであればバッテリー残量が少なくなってきた場合に，一旦人間に与えられたタスクの遂行中断し，充電を行ったり，人間に警告して電力を供給することで機能停止を防ぎ，再びタスクを行うことが可能となる．

第４章　提案システム

本章では，バッテリー残量に合わせて行動選択を行う学習システムを提案する．第 1 節では，自己保存を考慮するタスクについて述べる．次に，第 2 節で人間に与えられたタスクと自己保存を考慮したタスクのバランス調整について述べ，第 3 節以降で提案システムの具体的な構造，学習方法などについて述べる．

4.1 ロボットに対する 2 種類のタスクの設定

本研究では，自己保存を考慮するロボットが考慮すべきタスクについて考える．ロボットは人間に与えられた仕事を遂行することが目的となるので，第一に考慮すべきタスクは，人間に与えられるタスクとなる．また，ロボットが自己保存するためには人間に与えられるタスクの遂行とは別にロボットが自己保存を考慮する行動を取らなければならない．従って，第二に自己保存について考慮するタスクを設定しなければならない．

人間がロボットに与えるタスクとは，部屋の掃除であったり，荷物の運搬であったりと人間がロボットに仕事を要求するものである．これに対して，自己保存を考慮するタスクでは，ロボットが電力不足による機能停止を免れるために，バッテリーの充電や静止状態での待機，人間への警告などを目的とするタスクである．この 2 つのタスクは独立しており，ロボットは 2 つのタスクを持つこととなる．

人間に与えられたタスクはロボットにとって外部から与えられる要因であるので，外部タスクと定義する．反対にロボットが自己保存を行うことはロボットからみて内部の状態であるバッテリー残量を考慮するするタスクであるので内部タスクと定義する（図 4.1）．

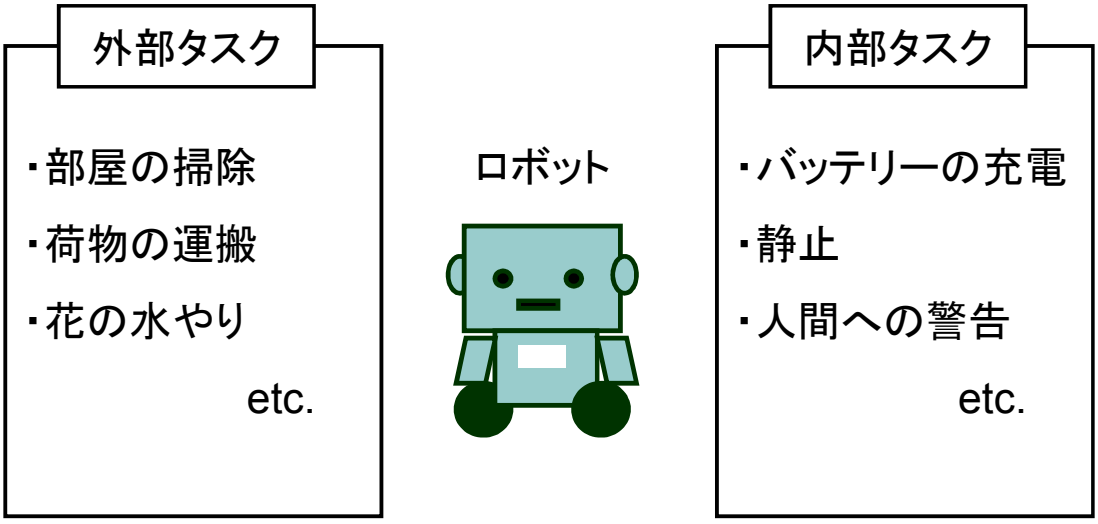


図 4.1　ロボットにおける外部タスクと内部タスク

4.2 外部タスクと内部タスクのバランス調整

前節では，ロボットにおける外部タスクと内部タスクについての定義を行った．しかし，ロボットに対して内部タスクを設定しただけではバッテリー残量に対する行動選択を行うことはできない．バッテリー残量に応じた行動選択を実現するためには，バッテリー残量に応じてどのような行動が適しているのかを考える必要がある．バッテリー残量が十分な場合は，人間に与えられたタスクの遂行が重要であり，機能停止を防ぐための自己保存は重要ではない．反対に，ロボットが外部タスクを遂行するために必要なバッテリー残量が十分でない場合には，人間に与えられたタスクの遂行よりも，機能停止を防ぐための自己保存が重要となる．この関係を先に定義した外部タスクと内部タスクに当てはめる．バッテリー残量が十分な場合には内部タスクの遂行よりも外部タスクの遂行を優先し，バッテリー残量が十分ではい場合には外部タスクの遂行よりも，内部タスクを優先すればよい．

本システムでは，バッテリー残量に応じて 2 つのタスクのうちどちらの行動をとれば良いのかというバランスを調整する（図 3.2）．バランスの調整方法については，いくつか考えられる．例えば，バッテリー残量が十分ある場合には，外部タスクの遂行を優先し，バッテリー残量が一定基準以下となった場合に内部タスクの遂行を優先するといった方法や，バッテリー残量毎にどちらのタスクを優先するか事前に設定する方法が挙げられる．しかし，このような方法では，外部タスクについての行動において消費されるバッテリーの量に応じて再設定が必要となる．そこで，本研究ではバッテリー残量毎に外部タスクと内部タスクのどちらを優先すべきか，ロボットが取る行動のバッテリー消費量に基づいてロボットが自律的に決定することを考える．ロボットが自律的に行動を選択するために強化学習によってロボットが取るべき行動を学習させ，バッテリー残量に応じた行動を獲得することを考える．

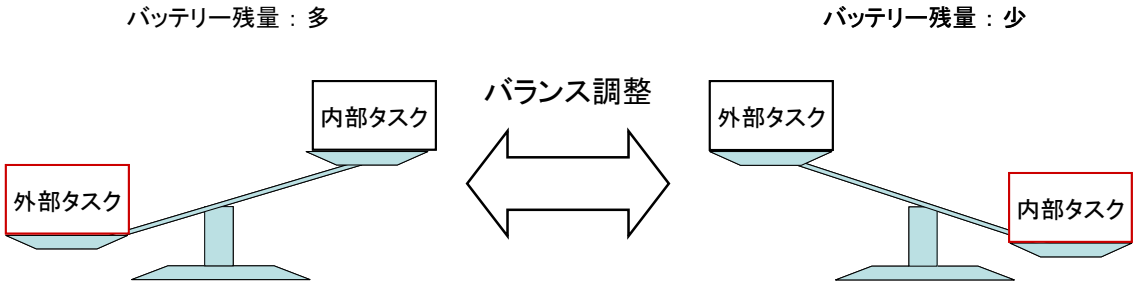


図 4.2　外部タスクと内部タスクのバランス調整のイメージ

4.3 提案システムの概要

本システムでは，バッテリー残量に対して外部タスクと内部タスクのどちらを優先すべきかを強化学習によって決定する．ロボットは外部タスクと内部タスクを持ち，双方のタスクにおいて最適な行動をそれぞれ独立して決定する．この結果，外部タスクを遂行するために最適な行動と，自己保存を考慮した場合最適な行動の 2 つの行動が得られる．得られた 2 つの行動とバッテリー残量から，バッテリー残量に対してどちらの行動が最適かロボットの行動におけるバッテリー消費量に基づいた強化学習によって学習し，行動選択を行う．ここで選択した行動が実際にロボットが最終的に行う行動となる．ロボットは行った行動に対して環境から報酬を獲得する．このサイクルを繰り返すことでバッテリー残量に合わせた行動を獲得する．

4.4 本システムの構成

本システムは 3 つのモジュールで構成されている．システムを構成するモジュールは外部タスク行動生成部，内部タスク行動生成部，バランス調整部となっている．提案する学習システムのシステム構成図を図 4.3 に示す. ロボットは外部タスクを遂行するために必要な行動を外部タスク行動生成部で獲得し，内部タスクを遂行するために必要な行動を内部タスク行動生成部で獲得する．最後に，バランス調整部において外部タスク行動生成部で生成された行動と内部タスク行動生成部で生成された行動，ロボットのバッテリー残量から，バッテリー残量に合わせた行動を強化学習によって獲得し，ロボットの行動を決定する．各モジュールの詳しい説明は以下の節で行う．

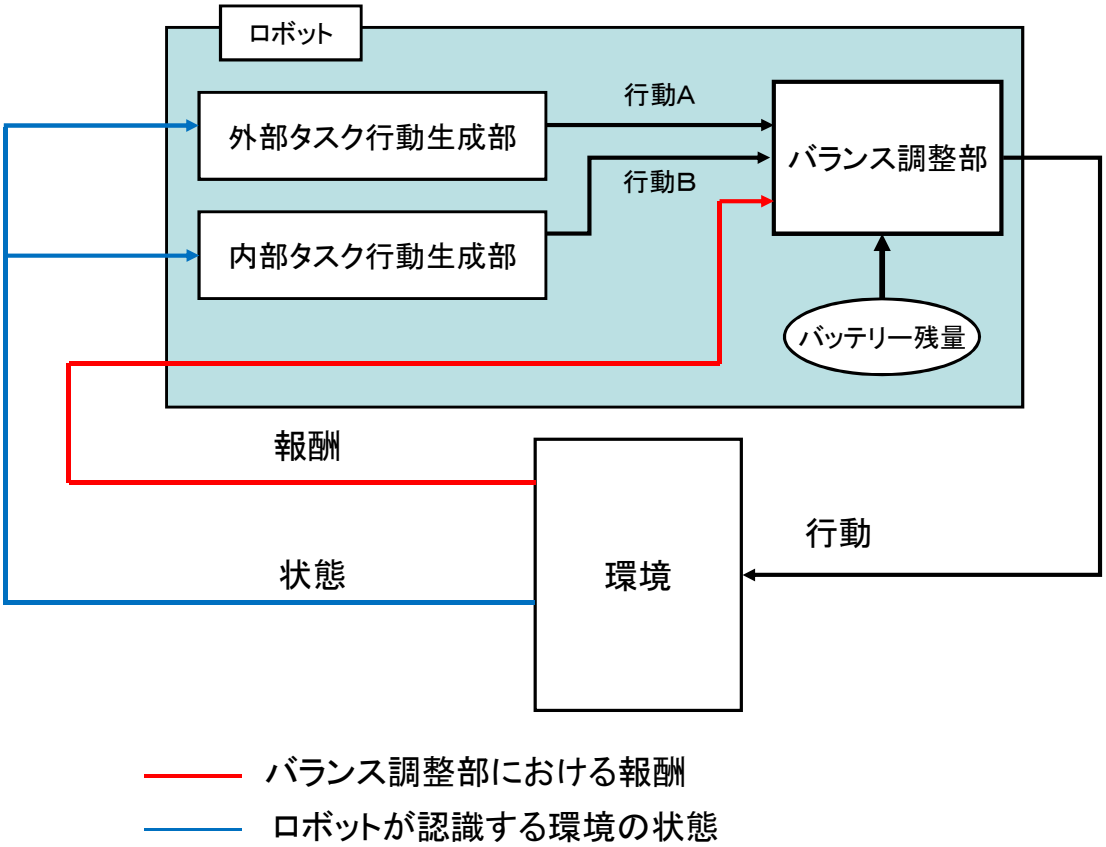


図 4.3　提案システムのシステム概念図

4.4.1 外部タスク行動生成部

外部タスク行動生成部は，ロボットが人間に与えられたタスクを遂行するために必要な行動を生成する部分である．ロボットが認識する環境の状態を入力とし，状態に応じて外部タスク遂行のための行動を出力する．生成された行動はバランス調整部へ入力される．

人間に与えられたタスクに必要な行動を学習によって獲得するロボットの場合には，外部タスク行動生成部に学習手法を適用することになる．学習の手法は人間がロボットに対して与えるタスクによって変化する可能性があるが，本研究では強化学習を適応したロボットを想定した場合の外部タスク行動生成部の構成について説明する．強化学習を想定すると，外部タスク行動生成部の入力はロボットが認識する状態と環境から獲得する報酬の 2 つとなる．また，出力は学習結果の行動となり，バランス調整部へ入力される．(図 4.4)

外部タスク行動生成部では，ロボットが認識した環境の状態と環境から得られる報酬から，状態行動対に対する行動価値を更新する．更新された行動価値から行動を選択する．外部タスクにおける強化学習の方法は第 2 章において説明した強化学習の学習方法と同様である．

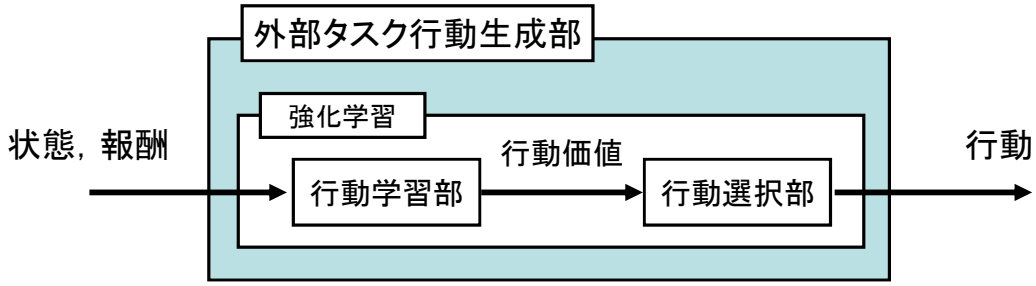


図 4.4　外部タスク行動生成部の入出力の関係（強化学習を適用した場合）

4.4.2 内部タスク行動生成部

内部タスク行動生成部は，ロボットが自己保存を行うために必要な行動を生成する部分である．ロボットが自己保存のために行う行動は，外部タスクの行動を行うために必要なバッテリー残量が不足した場合にロボットが取る行動となる．例えば，ロボットが静止状態となってバッテリー残量を維持する，バッテリーを充電するなどの行動が考えられる．ロボットが自己保存のために行う行動は，周りの環境の状態によって変化する可能性があるので，内部タスク行動生成部ではロボットが認識する環境の状態を入力とし，状態に応じて内部タスク遂行のための行動を出力する．生成された行動はバランス調整部へ入力される．

自己保存のための行動は，人間が事前に設定することも可能である．しかし，周囲の環境によって自己保存のための行動は変化することが考えられる．すぐに充電が可能な環境では，充電を行うことが自己保存のための行動となり，充電が不可能な場所では静止状態となり電力消費を抑えることが自己保存のための行動となり得る．このように，周囲の環境によって自己保存のための行動は変化する可能性があることから本研究では，内部タスクについても外部タスクと同様に強化学習を適用した場合を考え，内部タスク行動生成部に強化学習を適用した場合の構成について説明する．

内部タスク行動生成部の入力はロボットが認識する状態と環境から獲得する報酬の 2 つとなる．また，出力は学習結果の行動となり，バランス調整部へ入力される．（図 4.5）学習の方法は，外部タスク学習部と同様である．

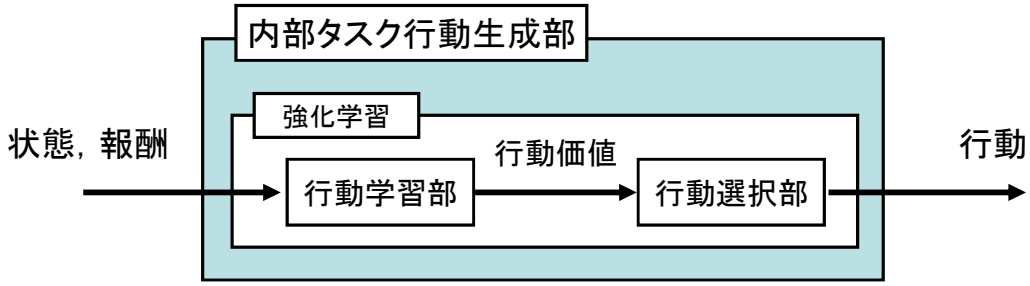


図 4.5　内部タスク行動生成部の入出力の関係（強化学習を適用した場合）

4.4.3 バランス調整部

バランス調整部は，バッテリー残量に対して外部タスクの遂行と内部タスクの遂行のどちらを優先するか決定する部分であり，本システムで最も重要なモジュールである．

バランス調整部は外部タスク行動生成部と内部タスク行動生成部それぞれにおいて生成される行動，バッテリー残量，ロボットが行った行動に対する報酬を入力とする．入力された各行動とバッテリー残量から，バッテリー残量とバッテリー消費量に合わせた行動を学習し，行動の選択を行う（図 4.6）．

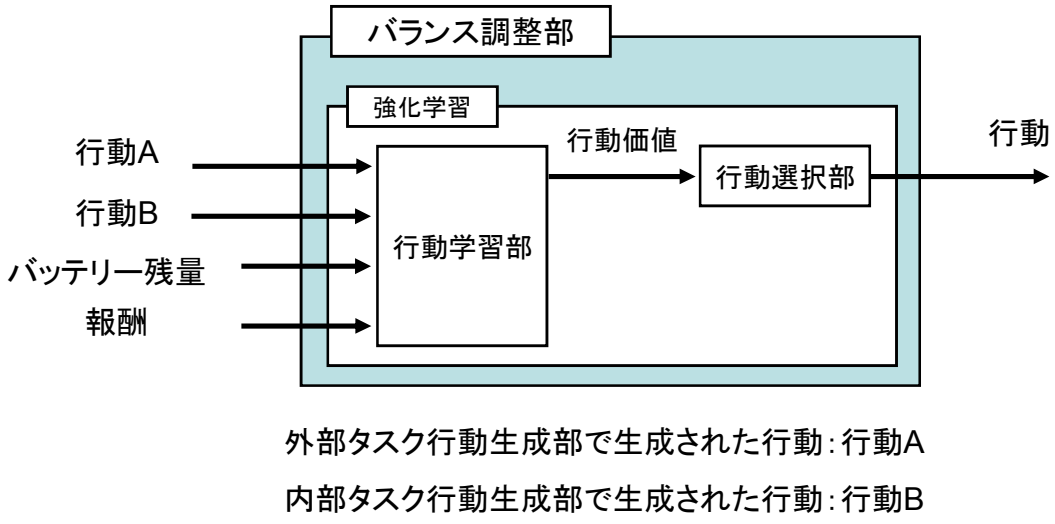


図 4.6　バランス調整部の入出力の関係

以下でバランス調整部の学習方法について説明する．バランス調整部での強化学習に用いる状態 s は，ロボットのバッテリー残量とする．バッテリー残量は満充電された状態をバッテリー残量 100%とし，バッテリー残量が最も多い状態を 100%，最も少ない状態を 0%とする．また，状態 s は，バッテリー残量 0%から 100%まで 1%刻みで設定し，状態数は 101 個となる．状態 s を式（4.1）に示す．

$$s = \{0,1,2,\ldots,100\} \tag{4.1}$$

状態sにおいてロボットが選択可能な行動aは，外部タスク行動生成部において生成された行動Aと内部タスク行動生成部において生成された行動Bのどちらかとなる．行動aを式（4.2）に示す．

$$a = \{a_1, a_2\} \tag{4.2}$$

a_1：外部タスク行動生成部で生成された行動A

a_2：内部タスク行動生成部で生成された行動B

バランス調整部では，上記で設定した状態sと行動aの組み合わせに対して，報酬を与え行動価値を更新する．

次に，バランス調整部で与える報酬について説明する．バッテリー残量が外部タスクの行動を取るために十分である場合，ロボットは人間に与えられた仕事である外部タスクの行動をとることが望ましい．逆にバッテリー残量が十分でない場合には，ロボットの自己保存のために内部タスクの行動を選択する．これは，バッテリー残量が多い場合には，外部タスクの行動に対する重要性が高く，内部タスクの行動に対する重要性が低いということを表わしている．反対に，バッテリー残量が少ない場合には，外部タスクの行動に対する重要性が低く，内部タスクの行動に対する重要性が高くなることを表わしている．そこで，外部タスクについての行動と内部タスクについての行動にそれぞれバッテリー残量に対する重要度を設定する．重要度とは，バッテリー残量に対する行動の重要性を表わす度合いである．バッテリー残量が十分な場合には，外部タスクについての行動の重要度は大きくなり，バッテリー残量が十分でない場合には小さくなる．反対に，バッテリー残量が十分な場合には，内部タスクについての行動の重要度は小さくなり，バッテリー残量が十分でない場合には大きくなる．バッテリー残量に対する外部タスクの重要度をw_e，バッテリー残量に対する内部タスクの重要度をw_iとした場合に，重要度を算出する式をそれぞれ式（4.3），（4.4）に示す．

$$w_e = \frac{1}{1+e^{\beta(d-v_b)}} \tag{4.3}$$

$$w_i = \frac{1}{1+e^{\beta(-d+v_b)}} \tag{4.4}$$

β, d は定数，v_b はバッテリー残量を表わしている．

$\beta = 0.2, d = 50$ とした場合のグラフを図 4.7，図 4.8 に示す．

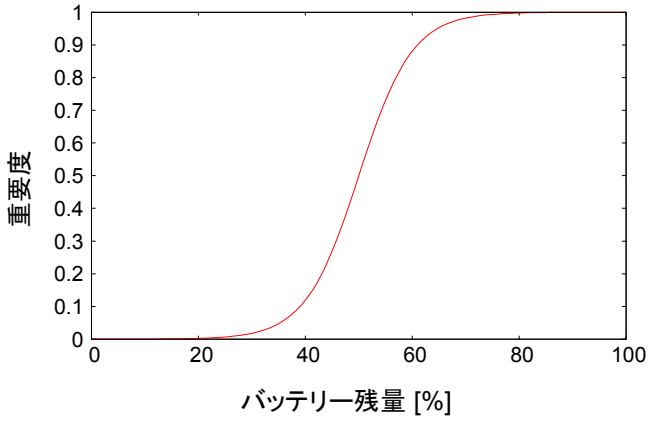


図 4.8　バッテリー残量に対する外部タスクの重要度

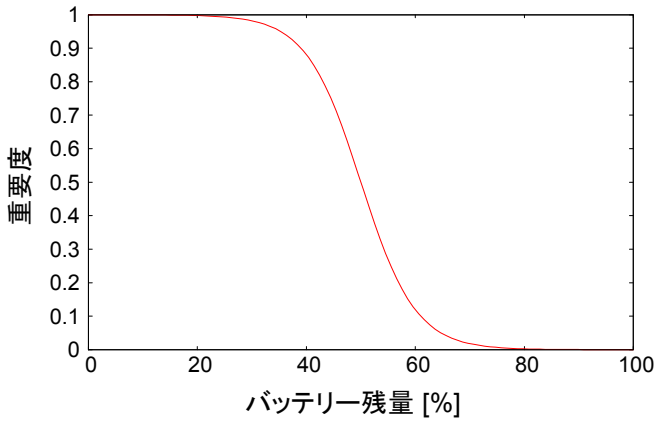


図 4.9　バッテリー残量に対する内部タスクの重要度

上記の重要度を用いて報酬関数を設定する．ある状態sで行動aを行った場合に，状態行動対(s,a)に与えられる報酬rを式（4.5），(4.6)に示す．

$$r = w_e r_e + w_i r_i \tag{4.5}$$

$$r_i = e^{(\frac{-v_c}{v_b})} \tag{4.6}$$

v_cとは，ある状態sに対してロボットが取った行動aのバッテリー消費量を表わしている．またv_bは，ロボットが行動aを行った場合のバッテリー残量を表わしている．

r_eはロボットがとった行動に対して人間が与える報酬である．ずなわち，外部タスクの行動を取った場合に高い報酬となり，内部タスクの行動を取った場合には低い報酬となる．また，外部タスク行動生成部において，強化学習を適用した場合には，外部タスク行動部において用いられる報酬関数をそのまま用いることになる．

r_iは，バッテリー残量が少ない状態で，バッテリー消費量の小さい行動を行うほど高い報酬となっている．さらに，r_eとr_iに対して，重要度を重みとして付加している． バッテリー残量が多い場合にはr_eに対する重みが増加し，r_iに対する重みが減少するので，外部タスクを遂行するための行動に対する行動価値が増加する．反対に，バッテリー残量が少なくなった場合には r_eに対する重みが減少し，r_iに対する重みが増加するので内部タスクを遂行するための行動に対する行動価値が増加する．式（4.5），式（4.6）の報酬関数に従い，状態行動対に対する行動価値を更新する． 更新された行動価値に応じてロボットの行動を選択する．

第 5 章 実験

5.1 実験の目的と概要

本実験は，シミュレーションで行う．本研究の目的としては２つあげられる．まず，提案する学習システムによって，ロボットがバッテリー残量に応じた行動を取ることが可能かどうかを検証する．また，バッテリー消費量の異なる外部タスクの行動に対し，バッテリー消費量に合わせた行動を取ることが可能かどうかを検証する． ロボットには外部タスクの行動と内部タスクの行動をあらかじめ設定し，バランス調整部へ入力する．入力された外部タスクと内部タスクの行動とバッテリー残量をもとにバランス調整部で学習を行い結果を検証する．

5.2 実験設定

本実験では，バッテリーによって電力を供給し，行動するロボットを対象とする．バッテリーはロボットが行動を取ることにより，その行動に応じてバッテリーを消費する．このロボットに提案システムを適用しシミュレーション実験を行う．

シミュレーション実験では，外部タスク行動生成部からバランス調整部へ入力される行動と，内部タスク行動生成部からバランス調整部へ入力される行動を仮想的に設定する．今回のシミュレーション実験では外部タスクと内部タスクの行動の生成には強化学習を適用しない．理由としては，外部タスク行動生成部と内部タスク行動生成部で実際にタスクを与えて学習を行った場合，出力された結果がバランス調整部での学習によって得られた行動であるのか，外部タスクについての学習や，内部タスクについての学習によって得られた行動であるのかが判断できなくなる可能性があるためである．今回のシミュレーション実験ではバランス調整部での学習結果のみの検証を行いたいので，外部タスク，内部タスクについての学習は適用しないこととする．

まず，外部タスク行動生成部で生成される行動を設定する．外部タスク行動生成部からバランス調整部へ入力される行動は 3 種類設定する．この 3 種類の行動をそれぞれ行動 A，行動 B，行動 C とし，各行動で消費するバッテリーの量が異なる．各行動に対するバッテリー消費量を表 5.1 に示す．

表 5.1 外部タスクの行動に対するバッテリー消費量

外部タスクの行動	バッテリー消費量
行動 A	20%
行動 B	10%
行動 C	6%

また，バランス調整部の報酬で用いられる r_e を表 5.2 のように設定する．この報酬の値はロボットの行動がどれだけ外部タスクの遂行に適している行動かを表わしている．従って，ロボットが外部タスクの行動を選択した場合に報酬は高くなり，外部タスクについての行動以外の行道すなわち内部タスクについての行動を行った場合には低い報酬となる．

表 5.2　外部タスクに対する報酬

行動	報酬値
外部タスクの行動が選択された場合の報酬	9
外部タスクの行動以外の行動が 選択された場合の報酬	2

次に，内部タスク行動生成部で生成される行動を設定する．内部タスク行動生成部からバランス調整部へ入力される行動は，行動 D としロボットは静止状態とする．静止状態であってもバッテリーは消費するが，外部タスクについての行動に対して少ない消費量を設定する．内部タスクについての行動を表 5.3 に示す．内部タスクの行動としては，充電や人間に警告するなどの行動も考えられるが，今回は最も単純なロボットの静止を設定する．

表 5.3 内部タスクの行動に対するバッテリー消費量

内部タスクの行動	バッテリー消費量
行動 D	2%

シミュレーション実験は，ロボットのバッテリー残量 100%の状態から開始し，バッテリー残量に対して外部タスクの行動と内部タスクの行動のどちらを選択すべきかを学習させる．ロボットが行動した結果バッテリー残量が 0%となった場合には，バッテリー残量は 100%まで回復し，再び行動を選択する．バッテリー残量が 0%となった場合には充電を行うためにロボットは行動すると考え，ロボットの行動回数を 1 回追加する．つまり，学習回数と行動回数は必ずしも一致しない．このサイクルを繰り返すことでバッテリー残量に対して最適な行動を学習させる．

シミュレーション実験としては 2 つの実験を行う．1 つ目の実験としてはバッテリー消費量の異なる外部タスクの行動A，B，Cそれぞれについて別々にシミュレーション実験を行い結果を検証する．

まず，外部タスクの行動を行動Aとし，内部タスクの行動を行動Dとした場合のシミュレーション実験の結果から検証を行う．続いて，外部タスクの行動Bと内部タスクの

行動Dとした場合のシミュレーション実験の結果から検証を行う．続いて，外部タスクの行動Cと内部タスクの行動Dとした場合のシミュレーション実験の結果から検証を行う．学習回数はそれぞれ 30000 回行う（図 5.1）． この場合の実験パラメータを表 5.4 に示す.

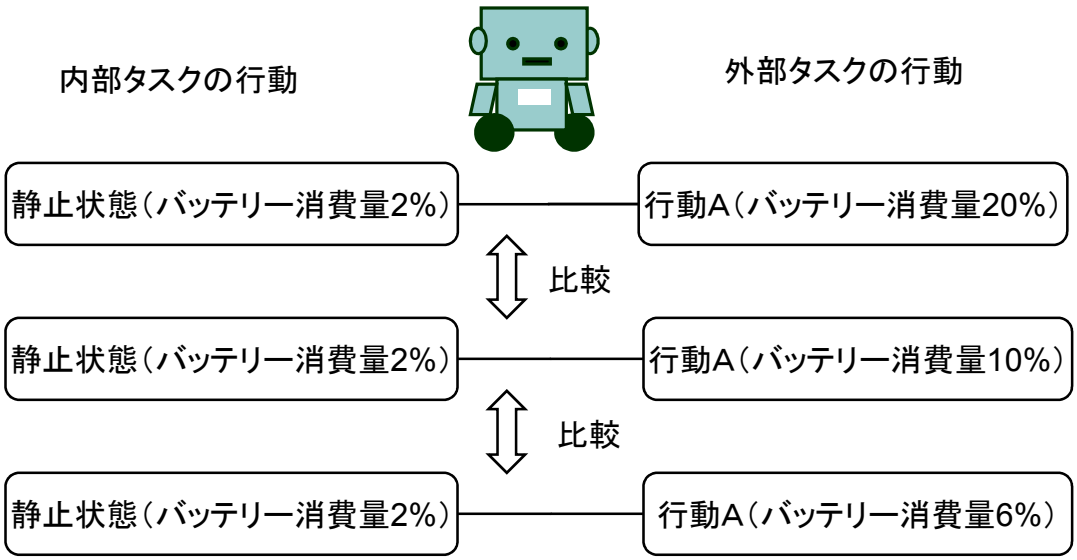


図 5.1　実験 1 の概念図

表 5.4 実験設定 1

学習回数	30000 回
初期行動価値	0.0
行動選択手法	ε -greedy 法
行動学習法	加重平均法
ε （ε -greedy 法）	0.1
ステップサイズパラメータ（加重平均法）	0.5
d （重要度のパラメータ）	50
β （重要度のパラメータ）	0.2

2 つ目の実験は，1 度のシミュレーション中に外部タスクの行動を行動A，行動B，行動Cと順に変化させ，内部タスクの行動Dとのバランス調整の検証を行う．行動 A，行動 B，行動 C は行動回数 20000 回毎に，行動A，行動B，行動Cと順に変化させる（図 5.2）．実験に用いるパラメータなどは表 5.5 に示す.

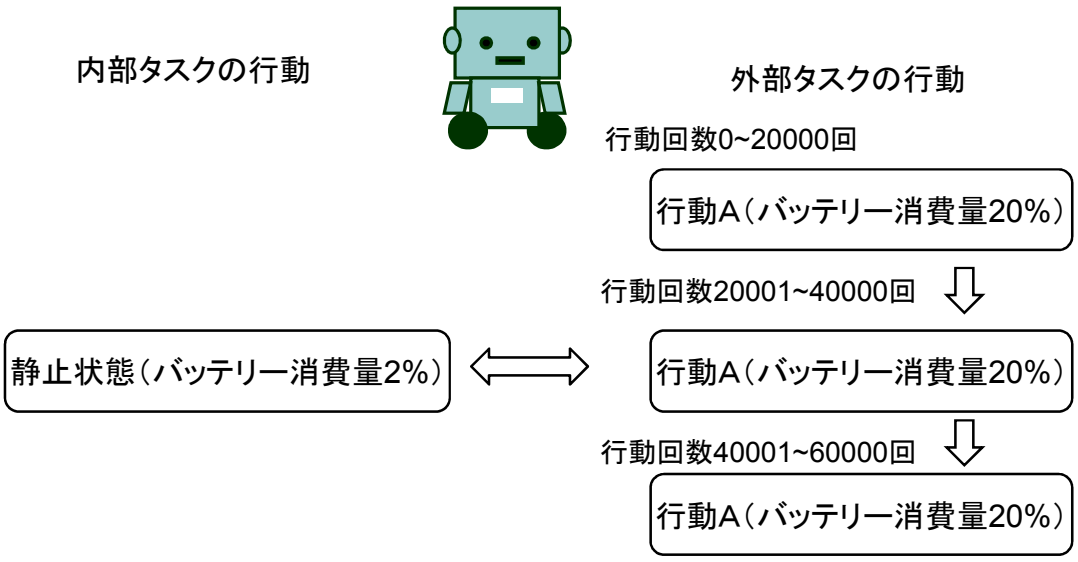


図 5.2　実験 2 の概念図

表 5.4 実験設定 2

学習回数	60000 回
初期行動価値	0.0
行動選択手法	ε -greedy 法
行動学習法	加重平均法
ε （ε -greedy 法）	0.1
ステップサイズパラメータ（加重平均法）	0.5
d （重要度のパラメータ）	50
β （重要度のパラメータ）	0.2

5.3 実験結果

5.3.1 外部タスクの行動Aに対する実験結果

外部タスクの行動を行動A，内部タスクの行動を行動Dとした場合のシミュレーション結果を示す．まず，学習初期のシミュレーション結果として行動回数 0 ~100 回に対するバッテリー残量の変化を図 5.1 に示す．また，行動回数に対するバッテリー消費量の変化を図 5.2 に示す．

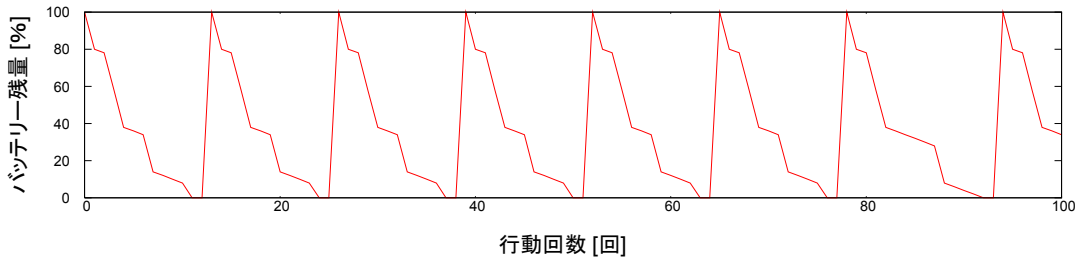


図 5.1 行動回数に対するバッテリー残量の変化（行動回数 0~100 回）

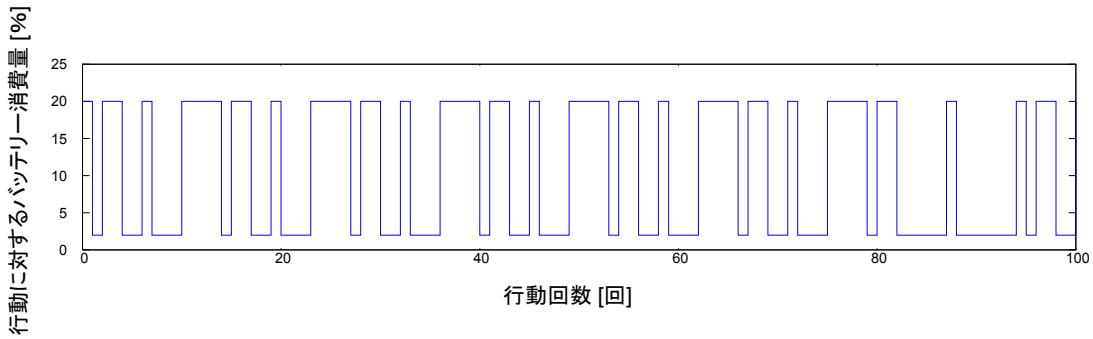


図 5.2 行動回数に対するバッテリー消費量の変化（行動回数 0~100 回）

次に，学習が収束したと思われる行動回数 20000 回から 20100 回における行動回数に対するバッテリー残量の変化を図 5.3 に示す．また，行動回数に対するバッテリー消費量の変化を図 5.4 に示す．また，図 5.3．図 5.4 を重ねたものを図 5.5 に示す．

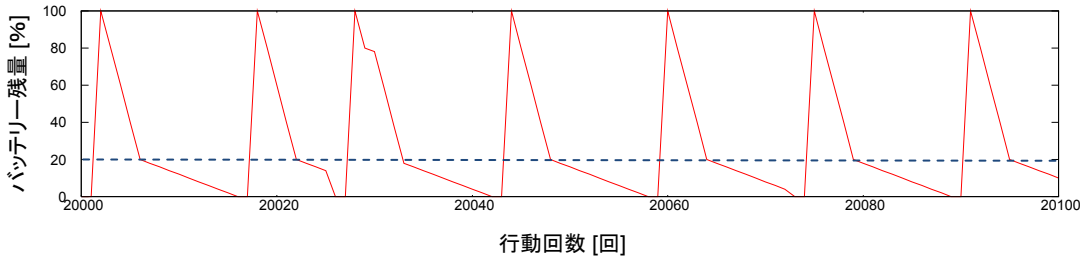


図 5.3 行動回数に対するバッテリー残量の変化（行動回数 20000~20100 回）

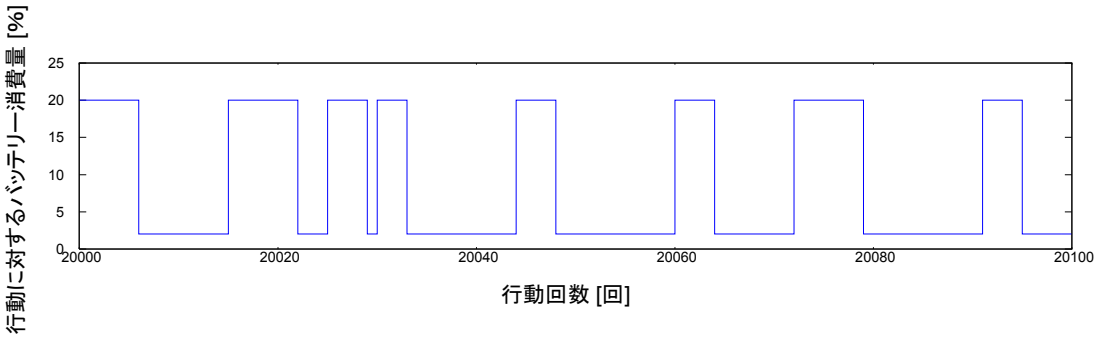


図 5.4 行動回数に対するバッテリー消費量の変化（行動回数 20000~20100 回）

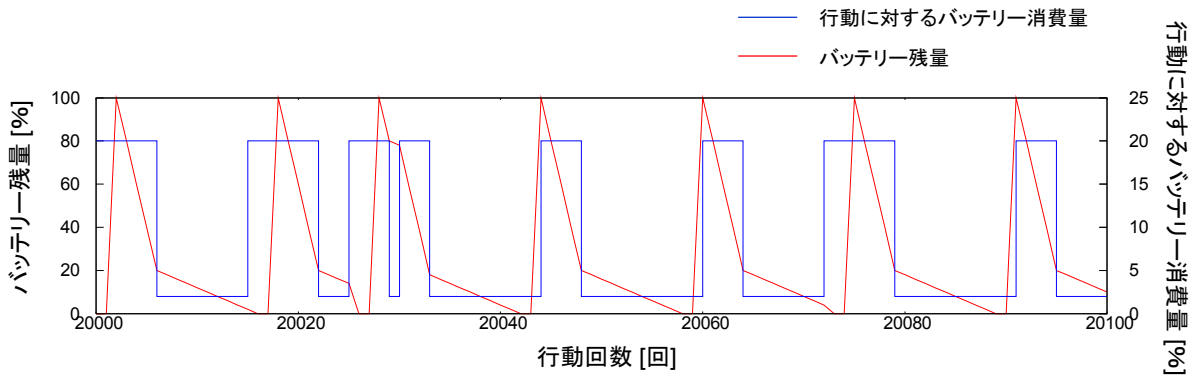


図 5.5　行動回数に対するバッテリー残量とバッテリー消費量の変化
（行動回数 20000~20100 回）

5.3.2 外部タスクの行動Bに対する実験結果

外部タスクの行動を行動B，内部タスクの行動を行動Dとした場合のシミュレーション結果を示す．まず，学習初期のシミュレーション結果として行動回数 0 ~100 回に対するバッテリー残量の変化を図 5.6 に示す．また，行動回数に対するバッテリー消費量の変化を図 5.7 に示す．

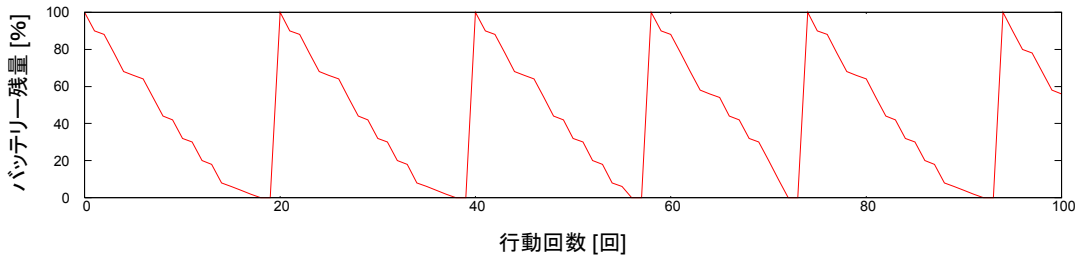


図 5.6 行動回数に対するバッテリー残量の変化（行動回数 0~100 回）

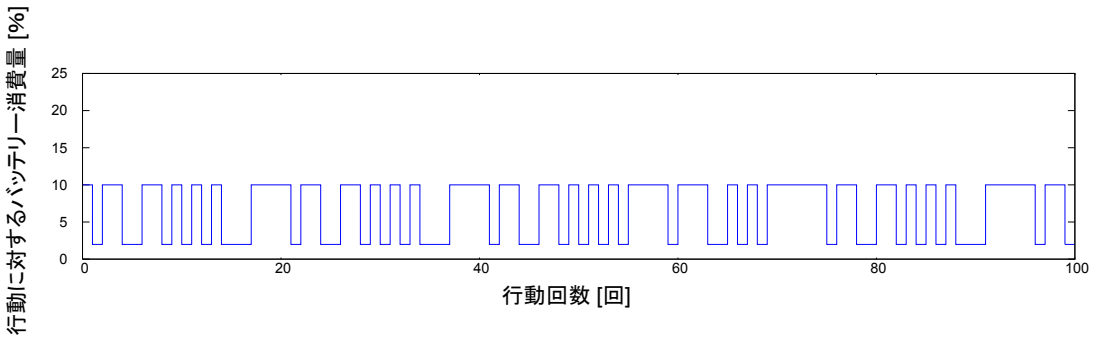


図 5.7 行動回数に対するバッテリー消費量の変化（行動回数 0~100 回）

次に，学習が収束したと思われる行動回数 20000 回から 20100 回における行動回数に対するバッテリー残量の変化を図 5.8 に示す．また，行動回数に対するバッテリー消費量の変化を図 5.9 に示す．また，図 5.8．図 5.9 を重ねたものを図 5.10 に示す．

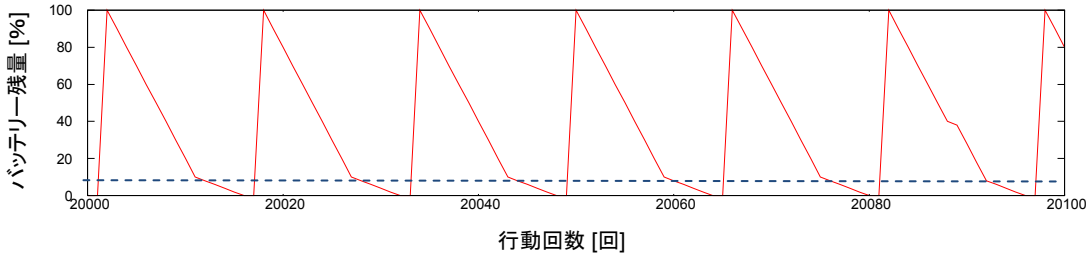


図 5.8 行動回数に対するバッテリー消費量の変化（行動回数 20000~20100 回）

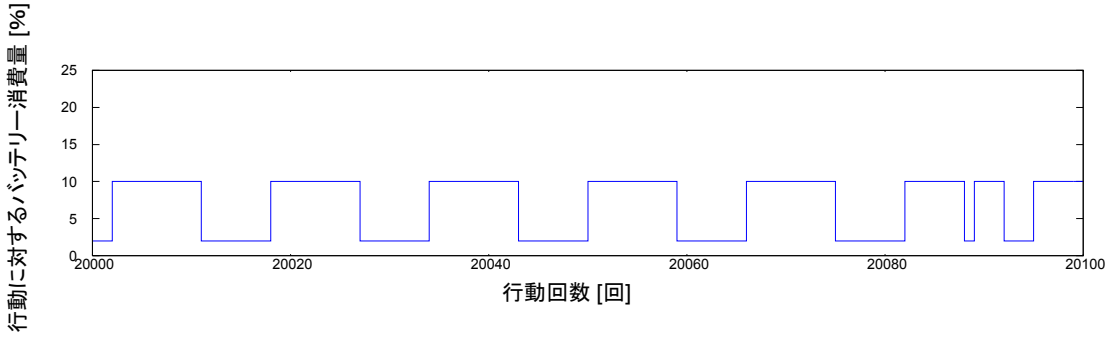


図 5.9 行動回数に対するバッテリー残量の変化（行動回数 20000~20100 回）

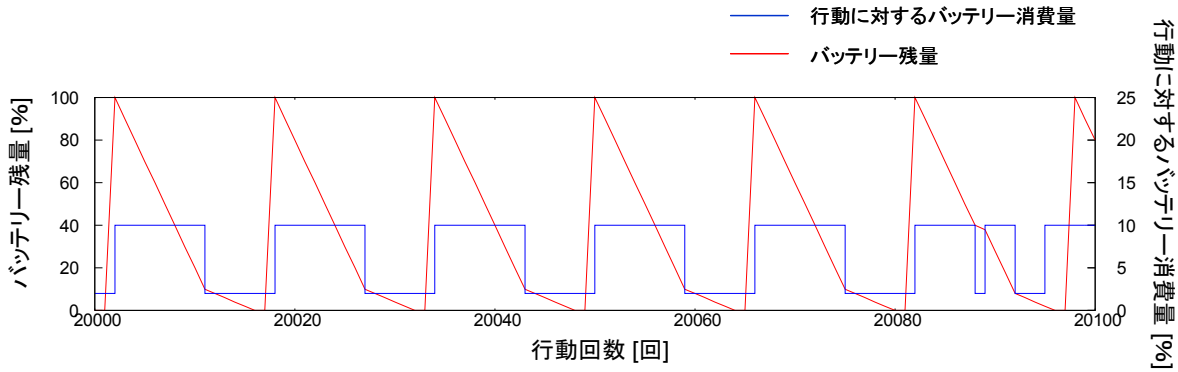


図 5.10　行動回数に対するバッテリー残量とバッテリー消費量の変化
（行動回数 20000~20100 回）

5.3.3 外部タスクの行動Cに対する実験結果

外部タスクの行動を行動C，内部タスクの行動を行動Dとした場合のシミュレーション結果を示す．まず，学習初期のシミュレーション結果として行動回数 0 ~100 回に対するバッテリー残量の変化を図 5.11 に示す．また，行動回数に対するバッテリー消費量の変化を図 5.12 に示す．

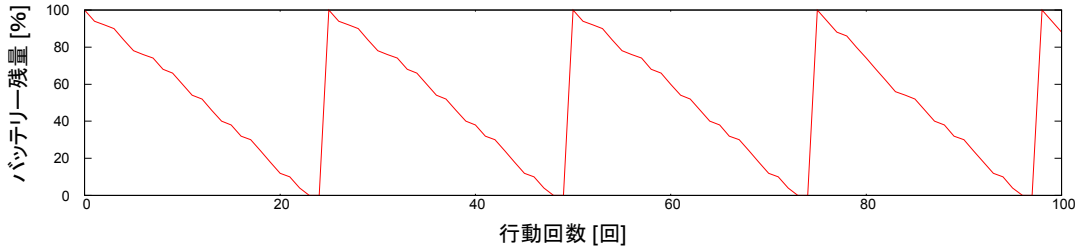


図 5.11 行動回数に対するバッテリー残量の変化（行動回数 0~100 回）

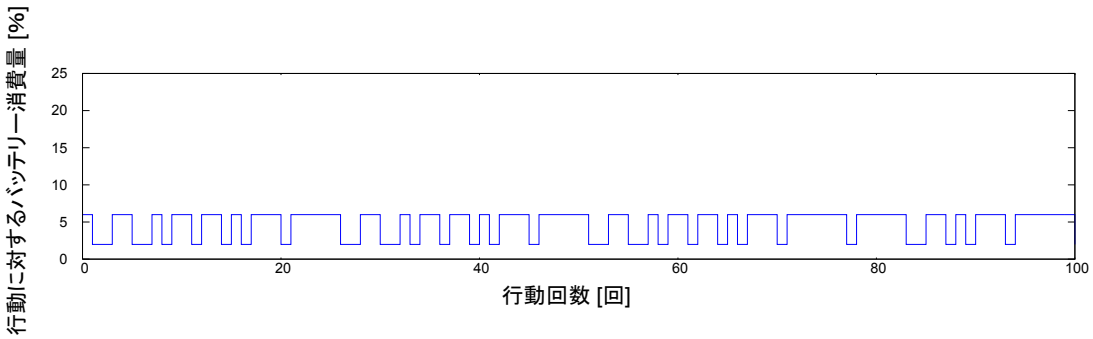


図 5.12 行動回数に対するバッテリー消費量の変化（行動回数 0~100 回）

次に，学習が収束したと思われる行動回数 20000 回から 20100 回における行動回数に対するバッテリー残量の変化を図 5.13 に示す．また，行動回数に対するバッテリー消費量の変化を図 5.14 に示す．また，図 5.13．図 5.14 を重ねたものを図 5.15 に示す．

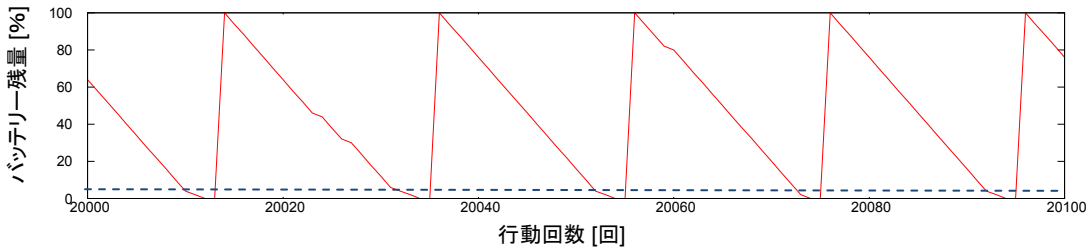


図 5.13 行動回数に対するバッテリー消費量の変化（行動回数 20000~20100 回）

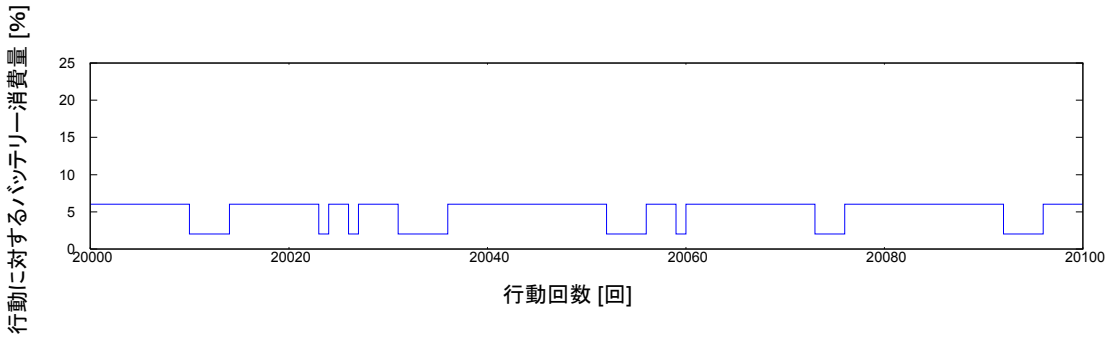


図 5.14 行動回数に対するバッテリー残量の変化（行動回数 20000~20100 回）

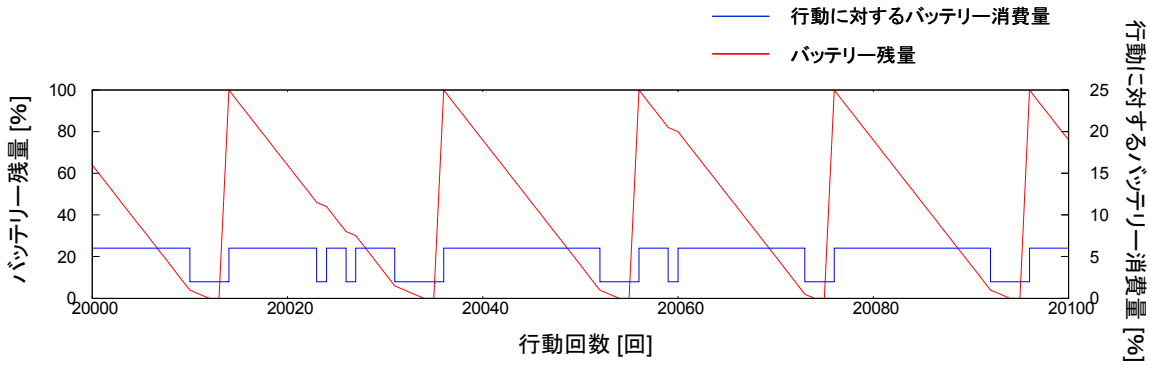
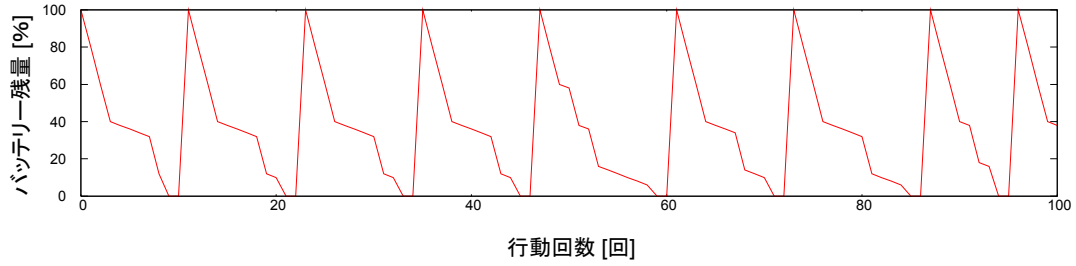


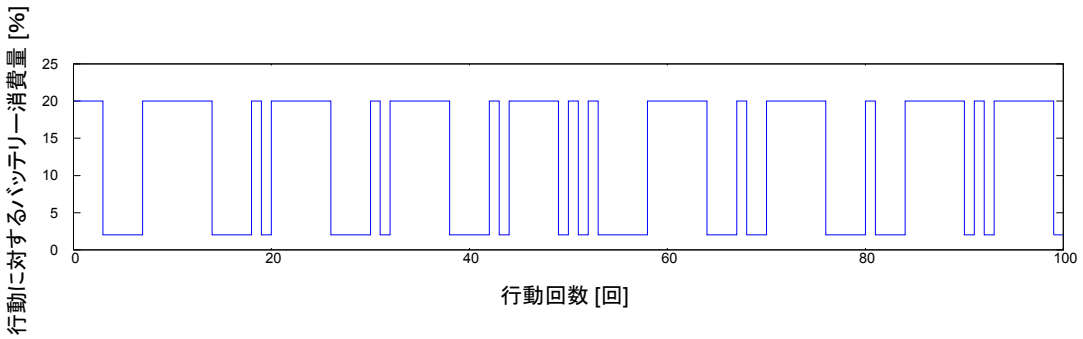
図 5.15　行動回数に対するバッテリー残量とバッテリー消費量の変化
（行動回数 20000~20100 回）

5.3.4 外部タスクの行動を順に変化させた場合の実験結果

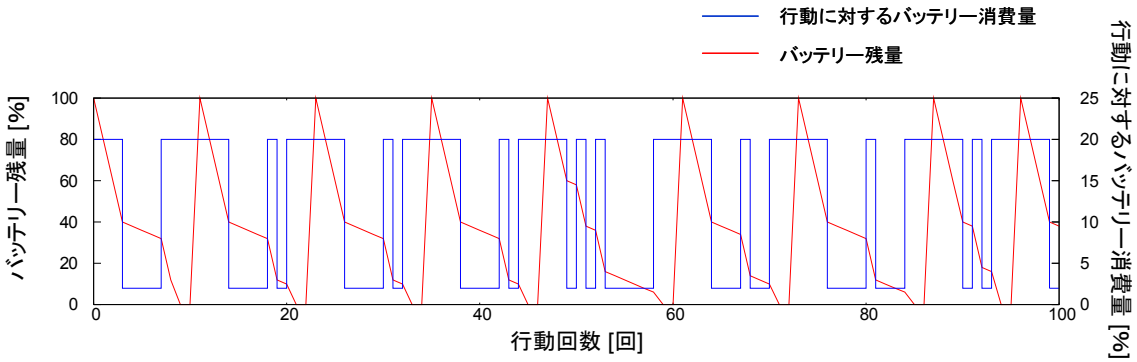
外部タスクの行動を行動 A，行動 B，行動 C の順で学習回数 20000 回毎に変化させた場合のシミュレーション結果を示す. 学習初期のシミュレーション結果として行動回数 0 から 100 回に対するバッテリー残量の変化と行動に対するバッテリー消費量の変化を図 5.16 に示す. また, 行動回数 20000~20100 回, 行動回数 25000~25100 回, 行動回数 40000～40100 回，行動回数 45000~45100 回，59000～59100 回時の行動回数に対するバッテリー残量の変化と行動に対するバッテリー消費量の変化をそれぞれ図 5.17，図 5.18，図 5.19，図 5.20，図 5.21 に示す.



（1）行動回数に対するバッテリー残量の変化

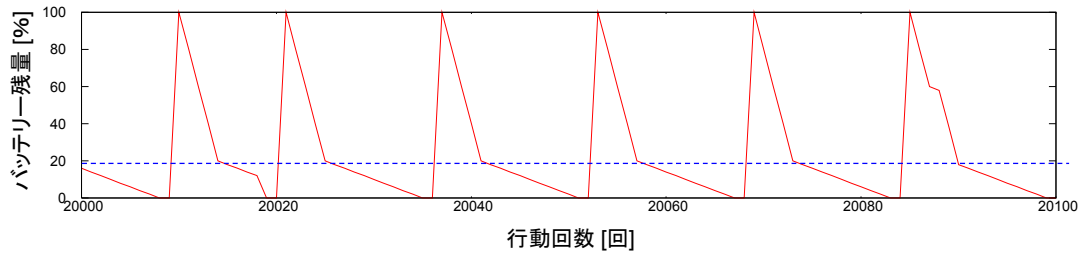


（2）行動回数に対するバッテリー消費量の変化

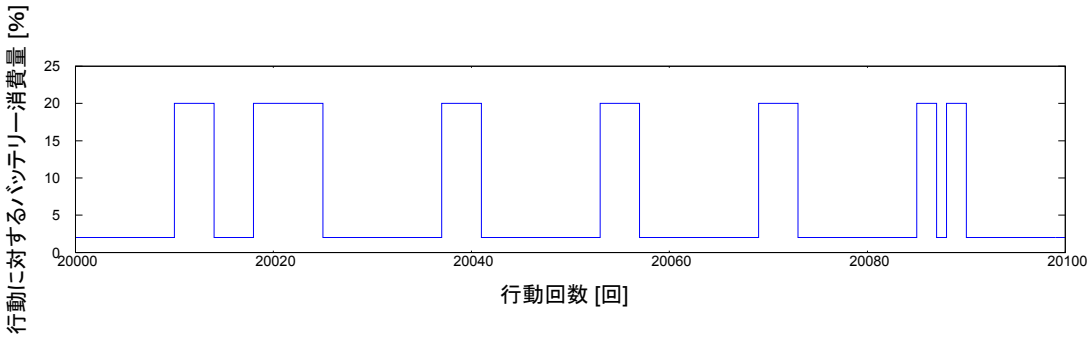


（3）バッテリー残量とバッテリー消費量を重ねた結果

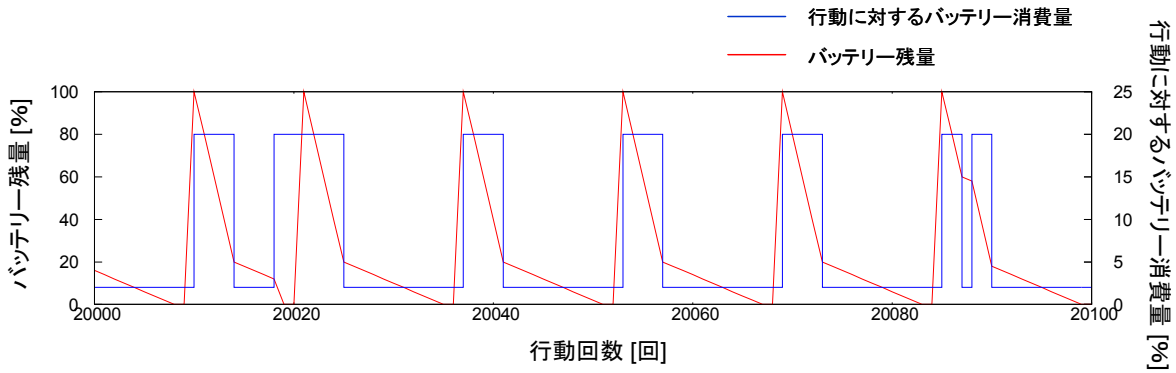
図 5.16　行動回数 0~100 回におけるシミュレーション結果



（1）行動回数に対するバッテリー残量の変化

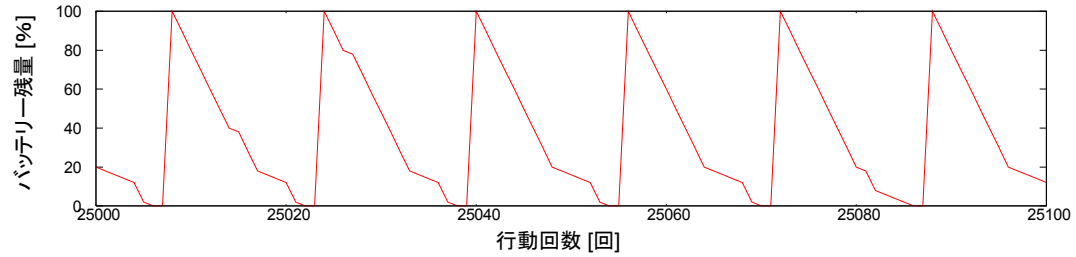


（2）行動回数に対するバッテリー消費量の変化

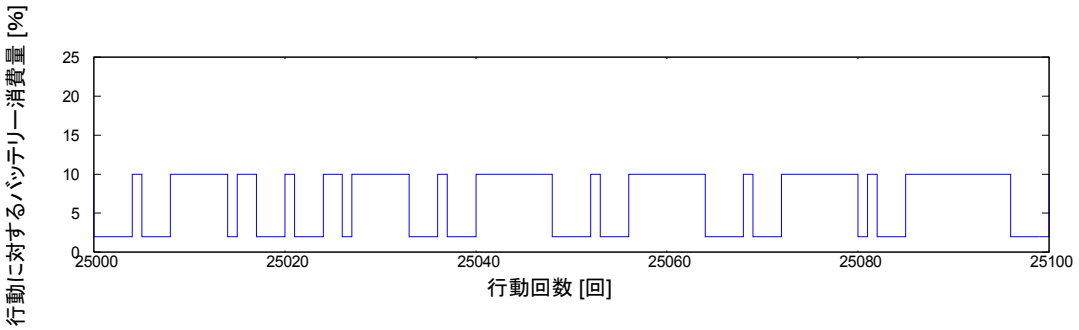


（3）バッテリー残量とバッテリー消費量を重ねた結果

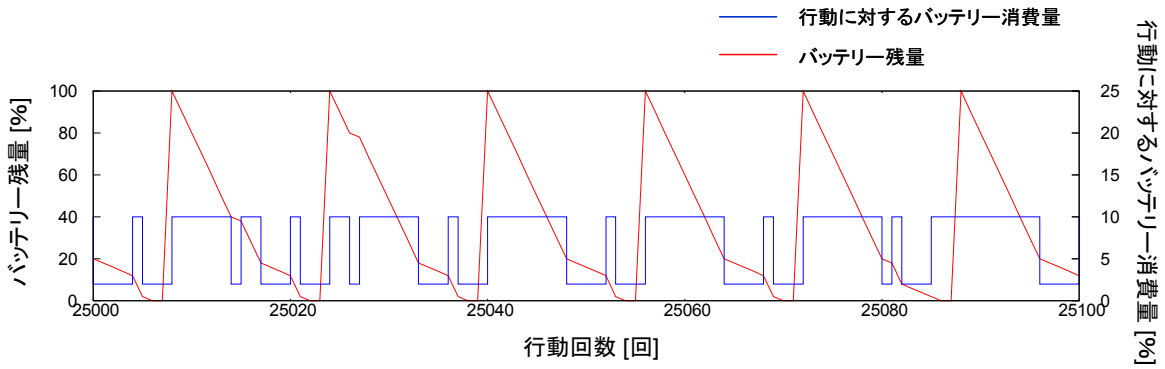
図 5.17　行動回数 20000～20100 回時のシミュレーション結果



（1）行動回数に対するバッテリー残量の変化

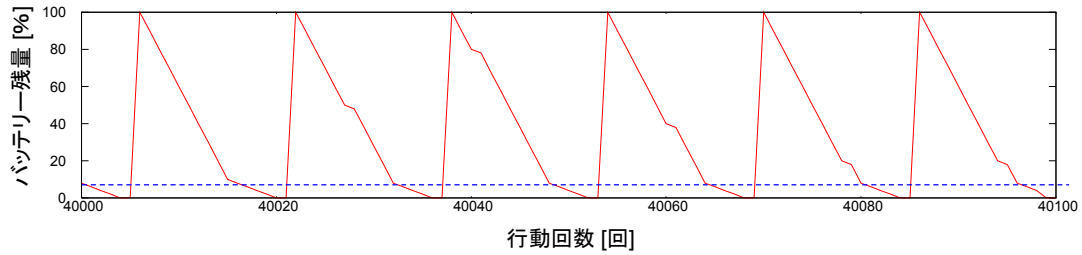


（2）行動回数に対するバッテリー消費量の変化

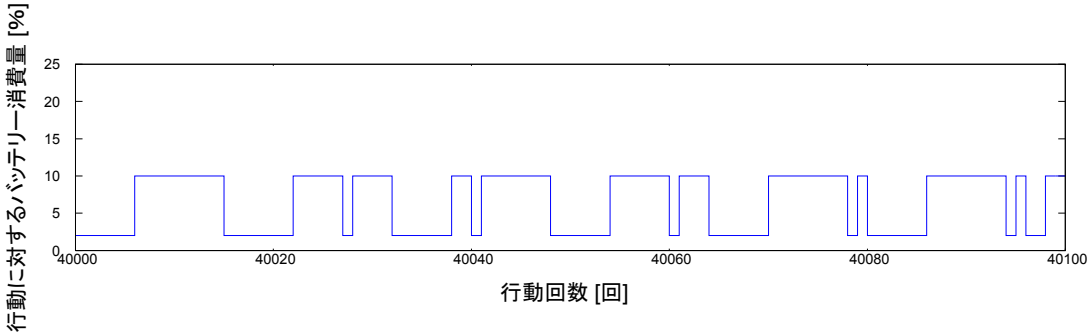


（3）バッテリー残量とバッテリー消費量を重ねた結果

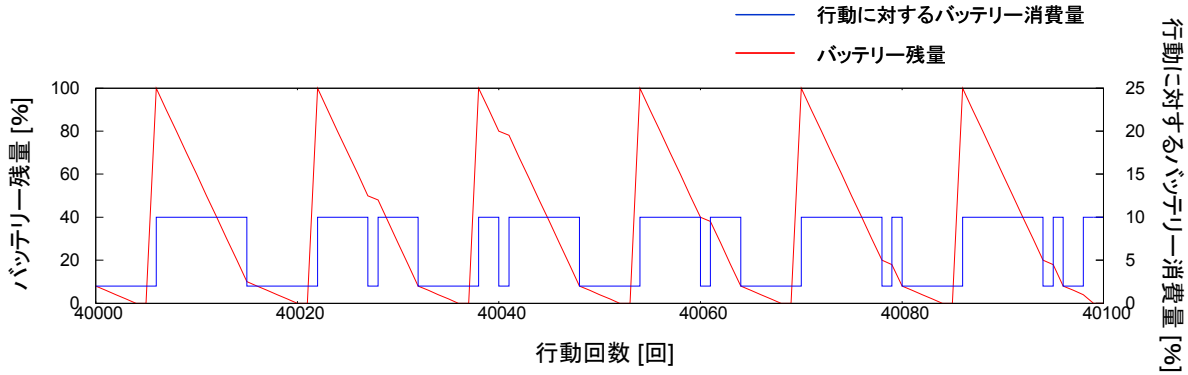
図 5.18　行動回数 25000～25100 回時のシミュレーション結果



（1）行動回数に対するバッテリー残量の変化

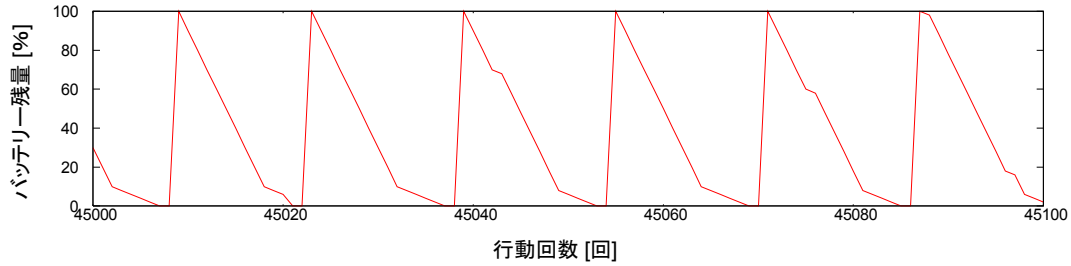


（2）行動回数に対するバッテリー消費量の変化

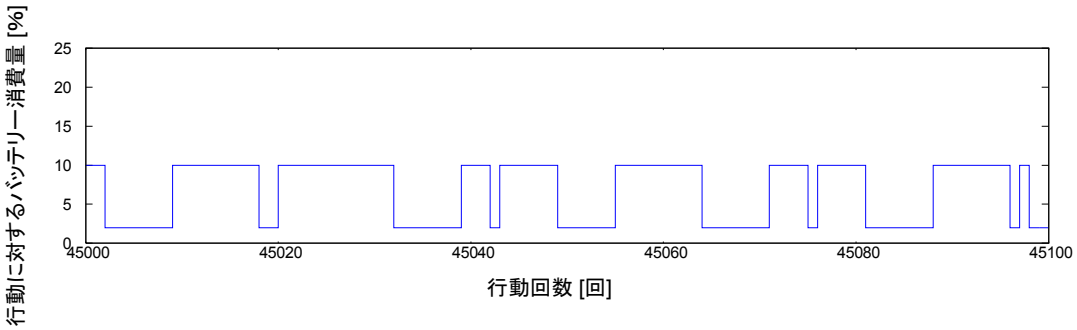


（3）バッテリー残量とバッテリー消費量を重ねた結果

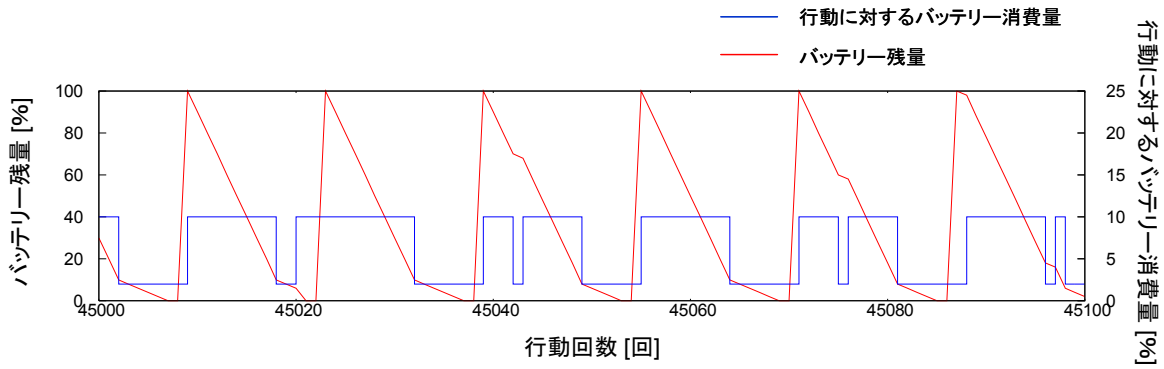
図 5.19　行動回数 40000～40100 回時のシミュレーション結果



（1）行動回数に対するバッテリー残量の変化

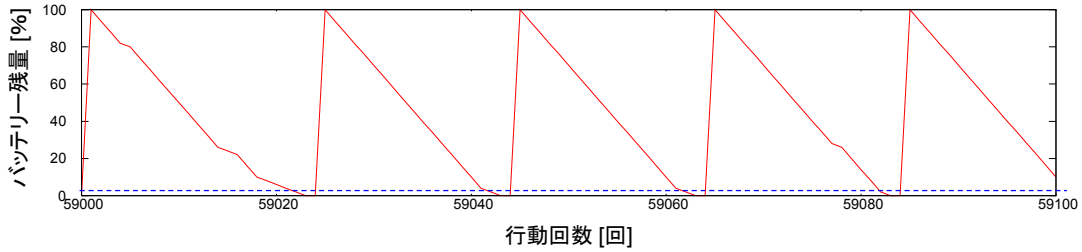


（2）行動回数に対するバッテリー消費量の変化

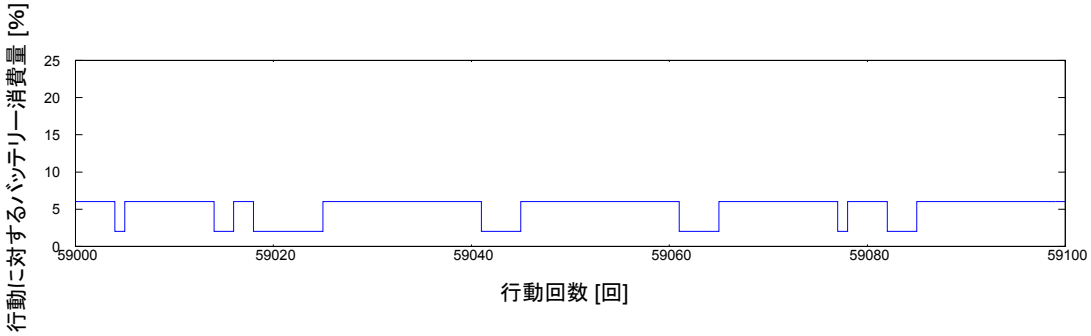


（3）バッテリー残量とバッテリー消費量を重ねた結果

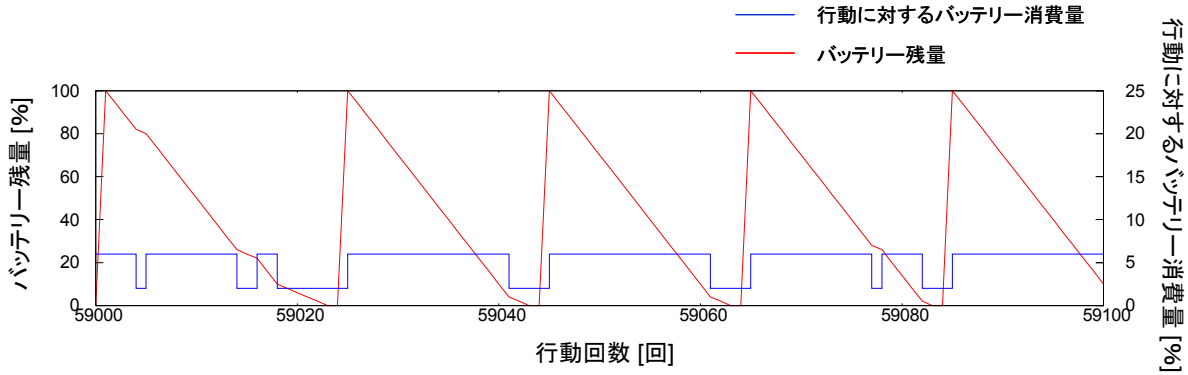
図 5.20　行動回数 45000～45100 回時のシミュレーション結果



（1）行動回数に対するバッテリー残量の変化



（2）行動回数に対するバッテリー消費量の変化



（3）バッテリー残量とバッテリー消費量を重ねた結果

図 5.21　行動回数 59000～59100 回時のシミュレーション結果

5.4 考察

5.4.1　外部タスクの行動Ａに対する実験結果の考察

行動回数に対するバッテリー消費量の変化は，各行動において外部タスクの行動と内部タスクの行動のどちらの行動を取ったかを表している．外部タスクの行動Aはバッテリー消費量が 20%であり，内部タスクの行動Dはバッテリー消費量が 2%である．従って，図 5.2 においてバッテリー消費量が 20%となっている部分では，ロボットは外部タスクの行動を取り，バッテリー消費量が 2%となっている部分では，内部タスクの行動を取っている．これを踏まえて図 5.1，図 5.2 をみると，学習初期の段階ではバッテリー残量に対して選択される行動はほぼランダムとなっている．選択される行動がランダムとなった原因としては，学習回数が少なく行動価値の更新が十分に行われなかったためだと考えられる．

次に図 5.3, 図 5.4, 図 5.5 より, 十分な学習を行った後の結果はバッテリー残量が 100%から 20%付近で外部タスクの行動を取り，バッテリー残量 20%から 0%までは内部タスクの行動を取るような傾向となっている．従って，バッテリー残量が 100%から 20%付近までロボットは人間に与えられた外部タスクを遂行し，20%付近で外部タスクの行動から内部タスクの行動へ切り替わり，バッテリー残量を消費が少ない静止行動を取っている．

また，結果においてバッテリー残量が多い場合でも内部タスクの行動を選択している箇所がある．これは，バランス調整部の行動選択手法に ε -greedy 法を用いているために，ε の確率でランダムな行動を選択したことが原因であると考えられる．

5.4.2　外部タスクの行動Ｂに対する実験結果の考察

行動回数に対するバッテリー消費量の変化は，各行動において外部タスクの行動と内部タスクの行動のどちらの行動を取ったかを表している．外部タスクの行動 B はバッテリー消費量が 10%であり，内部タスクの行動Dはバッテリー消費量が 2%である．従って，図 5.7 においてバッテリー消費量が 10%となっている部分では，ロボットは外部タスクの行動を取り，バッテリー消費量が 2%となっている部分では，内部タスクの行動を取っている．これを踏まえて図 5.6，図 5.7 をみると，学習初期の段階ではバッテリー残量に対して選択される行動はほぼランダムとなっている．選択される行動がランダムとなった原因としては，学習回数が少なく行動価値の更新が十分に行われなかったためだと考えられる．

次に図 5.8，図 5.9，図 5.10 より，十分な学習を行った後の結果はバッテリー残量が 100%から 10%付近で外部タスクの行動を取り，バッテリー残量 10%から 0%までは内

部タスクの行動を取るような傾向となっている．従って，バッテリー残量が 100%から10%付近までロボットは人間に与えられた外部タスクを遂行し，10%付近で外部タスクの行動から内部タスクの行動へ切り替わり，バッテリー残量を消費が少ない静止行動を取っている．

また，結果においてバッテリー残量が多い場合でも内部タスクの行動を選択している箇所がある．これは，バランス調整部の行動選択手法にε-greedy 法を用いているために，εの確率でランダムな行動を選択したことが原因であると考えられる．

5.4.3　外部タスクの行動Ｃに対する実験結果の考察

行動回数に対するバッテリー消費量の変化は，各行動において外部タスクの行動と内部タスクの行動のどちらの行動を取ったかを表している．外部タスクの行動 C はバッテリー消費量が 6%であり，内部タスクの行動Dはバッテリー消費量が 2%である．従って，図 5.12 においてバッテリー消費量が 6%となっている部分では，ロボットは外部タスクの行動を取り，バッテリー消費量が 2%となっている部分では，内部タスクの行動を取っている．これを踏まえて図 5.11，図 5.12 をみると，学習初期の段階ではバッテリー残量に対して選択される行動はほぼランダムとなっている．選択される行動がランダムとなった原因としては，学習回数が少なく行動価値の更新が十分に行われなかったためだと考えられる．

次に図 5.13，図 5.14，図 5.15 より，十分な学習を行った後の結果はバッテリー残量が100%から 6%付近で外部タスクの行動を取り，バッテリー残量 6%から 0%までは内部タスクの行動を取るような傾向となっている．従って，バッテリー残量が 100%から 6%付近までロボットは人間に与えられた外部タスクを遂行し，6%付近で外部タスクの行動から内部タスクの行動へ切り替わり，バッテリー残量を消費が少ない静止行動を取っている．

また，結果においてバッテリー残量が多い場合でも内部タスクの行動を選択している箇所がある．これは，バランス調整部の行動選択手法にε-greedy 法を用いているために，εの確率でランダムな行動を選択したことが原因であると考えられる．

5.4.4　外部タスクの行動を順に変化させた場合の実験結果の考察

図 5.16 は，行動回数 0~100 回のシミュレーション結果であり，学習初期でのシミュレーション結果を示している．この段階では，バッテリー残量に対して選択される行動はほぼランダムとなっている．選択される行動がランダムとなった原因としては，学習回数が少なく行動価値の更新が十分に行われなかったためだと考えられる．

図 5.17 は，行動回数 20000 から 20100 回時の実験結果である．学習回数 0 から 20000 回までは外部タスクの行動としては行動Ａが設定されている．行動Ａのバッテリー消費量は 20%である．実験結果よりバッテリー残量 20%付近で外部タスクの行動から内部タスクの行動へ切り替えることができている．

図 5.18 は，行動回数 25000 から 25100 回時の実験結果である．このとき外部タスクの行動は行動Ａから行動Ｂに変更されており，バッテリー残量 20%以下となった場合ではロボットが取る行動が安定していない．これは，ロボットの行動に対する行動価値が行動Ａに対する行動価値から行動Ｂに対する行動価値に合わせて最更新されているのが原因と考えられる．

図 5.19 は，行動回数 40000 から 40100 回時の実験結果である．このとき外部タスクの行動は行動Ｂであり，行動Ｂのバッテリー消費量は 10%である．実験結果よりバッテリー残量 10%付近で外部タスクの行動から内部タスクの行動へ切り替えることができている．

図 5.20 は，行動回数 45000 から 45100 回時の実験結果である．このとき外部タスクの行動は行動Ｂから行動Ｃに変更されている．しかし，外部タスクの行動が行動Ｂであったときの結果とほとんど変わらない結果となっている．これは，ロボットの行動に対する行動価値が行動Ｂに対する行動価値から行動Ｃに対する行動価値に合わせて最更新されている途中であるために行動Ｃに対する行動価値が獲得できていないためだと考えられる．

最後に図 5.19 は，行動回数 59000 から 59100 回時の実験結果である．このとき外部タスクの行動は行動Ｃであり，行動Ｃのバッテリー消費量は 6%である．実験結果よりバッテリー残量 6%付近で外部タスクの行動から内部タスクの行動へ切り替えることができている．

5.4.5　考察のまとめ

まず，外部タスクの行動 A，行動 B，行動 C の各行動に対するバランス調整についてまとめる．外部タスクの行動 A と内部タスクの行動 D とのバランス調整では，バッテリー残量 20%付近でロボットが取る行動は外部タスクの行動から内部タスクの行動へ切り替わっている. また, 外部タスクの行動 B と内部タスクの行動 D とのバランス調整では，バッテリー残量 10%付近で，ロボットが取る行動は外部タスクの行動から内部タスクの行動へと変化している．外部タスクの行動 C と内部タスクの行動 D とのバランス調整も同様に, バッテリー残量 6%付近でロボットが取る行動は外部タスクの行動から内部タスクの行動へと変化している．外部タスクの行動 A，行動 B，行動 C はそれぞれバッテリー消費量が異なる．従って，提案システムではバッテリー消費量の異なる外部タスクの行動に対して，各行動に応じたバランス調整を行うことが可能であることが示せた.

次に，ロボットにおける外部タスクの行動を行動 A，行動 B，行動 C の順に変化させた場合についてまとめる．ロボットの活動中に外部タスクの行動に対するバッテリー消費量が変化した場合でも，新たな行動に対する再学習を行うことでバッテリー消費量に合わせてバッテリー残量に応じた行動を獲得できることを示せた.

第６章 結論

6.1 全体を通してのまとめ

本研究では，ロボットがバッテリー不足による機能停止を回避するために自己保存機能が必要であることを示し，ロボットの自己保存機能としてバッテリー残量に合わせた行動選択を行う学習システムの実現を目標とした．そこで，本研究ではロボットに人間に与えられるタスクと自己保存を考慮するタスクをそれぞれ外部タスク，内部タスクと定義した．定義した外部タスクと内部タスクそれぞれの行動について，バッテリー残量に対してどちらの行動が適しているのかをロボットに学習させる．これにより，ロボットの自己保存としてバッテリー残量に合わせた行動選択を行わせるシステムを提案した．

また，提案システムによって，バッテリー残量に対して外部タスクの行動と内部タスクの行動とのバランス調整が行えるか検証するためにシミュレーションによる実験をおこなった．実験結果としては，バッテリー残量が多い場合には，外部タスクの行動を優先し，バッテリー残量が少なくなった場合には，外部タスクの行動から内部タスクの行動への切り替えが可能であることを示した．また，バッテリー消費量の異なる外部タスクの行動に対して，バッテリー消費量に応じて外部タスクの行動から内部タスクの行動へり変るバッテリー残量が変化することを示した．以上のことから，本研究ではロボットの行動に対するバッテリー消費量に基づき，バッテリー残量に対する行動を自律的に決定するロボットを実現することができた．

6.2 今後の課題

6.2.1 外部タスクと内部タスクについての学習

今回のシミュレーション実験では，外部タスク行動選択部と内部タスク行動選択部において学習の適用を行わず，事前に出力される行動を設定した．しかし，実際のロボットに本システムを適用するためには，人間に与えられたタスクを学習するロボットを想定しなければならない．よって，ロボットへ何からの外部タスクを与え，外部タスクの行動を学習させた上でバランス調整部へ学習結果である行動を入力し，今回のシミュレーション実験同様の結果が得られるかどうか検証を行う必要がある．同様に内部タスクにおいても環境に合わせた機能停止回避のための行動を学習させ，シミュレーション実験を行う必要がある．

6.2.2 バッテリーの充電を考慮したバランス調整

今回のシミュレーション実験では内部タスクの行動を静止状態と設定したため，実際のロボットに適応しても機能停止までの時間を延ばすことしかできない．しかし，内部タスクの行動に充電を設定すればロボットは機能停止を防ぐことが可能となり，長期的に活動することができると考えられる．従って，内部タスクの行動にロボットの充電を設定し，バッテリー残量が回復することも考慮したシミュレーション実験をおこない，充電を設定した場合でも提案システムが正しく機能するのかどうかを検証する必要がある．

6.2.3 実機実験

本研究では，シミュレーションでの実験しか行っていない．そのため，実際のロボットに提案システムを適用した場合に提案システムが有効に機能するかどうかはわからない．また，実機に提案システムを適用する場合には，適用するロボットの形態や搭載するバッテリーの種類など考慮すべき問題が多々あると考えられる．しかし，最終的な目標としては実機に提案システムを適用し提案システムの有用性や実機へ適用した場合の問題点などを検証する必要がある．

謝辞

本論文を結ぶにあたり，日頃より懇切なるご指導を賜りました倉重健太郎先生に深く感謝の意を表します．また，ご指導，ご助言をいただいた畑中雅彦先生，本田泰先生，佐賀聡人先生に感謝の意を表します．そして，論文の査読や助言をしていただいた認知ロボティクス研究室の木島康隆さん，中南義典さん，宮崎愛央さん，梅津祐介さん，北山直樹さん，渋谷和さん，杉本大志さん，高泉昇太郎君，沼田利伸君に感謝いたします．

参考文献

[1] 浅田 稔，野田 彰一，俵積田 健，細田 耕“視覚に基づく強化学習によるロボットの行動獲得” Vol.13, No.1, pp.68-74，1995

[2] 齋藤 史倫，福田 敏男“強化学習による実ロボットの運動制御”，Vol.13，No.1，pp.82-88, 1995

[3] 山田 朋史，永谷 圭司，田中 豊 “自律移動ロボットの自己保存-自律移動，6 軸力覚センサを用いたコンセント挿入動作の実現-”第 5 回システムインテグレーション部門学術講演会，2004，pp.725-726

[4] 八木 迪子，ピトヨハルトノ，鈴木 健嗣，橋本 周司 “エネルギー自給型屋外環境ロボット” 電子情報通信学会講演論文集 2004 年_情報・システム(1)，97，2004

[5] 大内 東“室内バルーンロボットにおける自律充電のドッキング制御”
情報処理学会研究報告．EC，エンタテインメントコンピューティング，Vol.2005，No.125，pp.1-5，2005

[6] 皆川 良弘，川村 秀憲，山本 雅人，高谷 敏彦，大内 東“充電地点へのドッキングを行う室内バルーンロボットの学習制御”，情報科学技術フォーラム一般講演論文集 Vol.5，No.3，pp.301-303，2006

[7] R.S. Sutton and A.G. Barto “Reinforcement learning An Introduction.”1998
邦訳：三上 貞芳，皆川 雅章 “強化学習”，森北出版，2000

[8] 森川 幸人“マッチ箱の AI”，新紀元社，2000

[9] 守田 観輝夫，石川 眞澄“強化学習を用いた生存欲に基づく行動の創発”
電子情報通信学会技術研究報告. NC, ニューロンコンピューティング, Vol.108, No.480, pp.279-283，2009

[10] 尾形 哲也，菅野 重樹，“自己保存に基づくロボットの行動生成 –方法論と機械モデルの実体化-”，日本ロボット学会誌，Vol.15，No.5，pp.710-721，1997