平成 18 年度

電気通信大学 大学院　修士論文

技能獲得過程の計算モデル

学　　籍　　番　　号　　0551027

氏　　名　　田口 林太郎

情 報 シ ス テ ム 学 科　　情報ネットワーク学専攻

指　　導　　教　　員　　阪口 豊　助教授

提　　出　　日　　平成 19 年 1 月 30 日

目次

第 1 章

序論

1.1 研究の背景および目的

ある運動を初めて行う初心者の運動とその運動に十分に習熟した上級者の運動とには大きな差がある．しかし，その差は初心者が学習を行うことによって次第に縮まっていく．この両者の運動の差は何に起因しているのか，また，初心者の学習が進行するにつれ何が変化し両者の差は縮まっていくのだろうか．

こうした運動学習をテーマにした研究は様々な分野で行われている．例えば，認知心理学の分野では学習の過程を観察し，モデリングを試みる研究が行われている[1]．しかし，こうした研究は抽象的なモデルの提示に留まり，そのモデル上で「技能差が何に起因しているのか」といった具体的な形での検証を行うことは難しい．一方，体育学やスポーツ科学といった分野では主に，上級者の運動の解析や運動学習における技能習熟の促進を目指した取り組み（コーチング）についての研究が行われている．また，研究というほどではないが，実際のスポーツの競技の場では効率の良い練習方法などが広く議論されている．これらはおおまかに見て，運動や学習の結果を論じるものであり，その時，学習者の中でどのような変化が起こっているのかということについては言及していない．

こうして考えると，一方は運動学習についての抽象的なアプローチ，もう一方は具体的なアプローチであり，これらの議論・研究にはしばしば乖離が見られるのが現状である．これは，それぞれの議論を横断的に考えられる土台となるものが未だないからである．

そこで本研究では，まず初級者と上級者の違い，そして初級者が上級者へと変化していく過程について考察し，運動学習について，実際のスポーツ競技の場，体育学やスポーツ科学，認知心理学的研究，シミュレーションモデル主体の神経科学的研究，でのそれぞれの議論を横断的に考えることのできる，理論モデル，シミュレーションモデルを構築することを目的とする．そして，そのモデル上で技能獲得過程の検証を行う．

第 2 章

運動学習

2.1 運動技能

一般に技能とは，身体を使ったある特定の作業を実行する過程において，多様な状況に適応して高度な精密さと安定したパフォーマンスの両方を実現する能力を伴った活動を意味している [2]．運動に関する技能をよりよいものへと向上させていくことが，運動学習の目的である．

運動技能には大別して二つの要素がある．一つは，筋力，柔軟性，敏捷性，持久性といった基礎的運動要因である．これらは，筋肉，関節，呼吸循環器系などの能力であり，生成された運動指令に基づき実際に運動を行う際に必要な能力である．

これに対するもう一つの要素は，運動すべき空間位置を記憶したり，運動パターンを記憶したりといった実際の運動と違いその行動が表出しない運動指令が生成されるまでの内的過程である．

運動技能の差が一つめの基礎的運動要因に起因するならば，練習を繰り返すうちに，筋力や持久性が向上しその差は縮まっていくだろう．しかし，我々は運動の上達がこの基礎的運動要因のみによらないことを経験的に知っている．ジャグリングや楽器演奏など，筋力などの要素によらない運動もある．

つまり，何が変化することによって運動が上達していくのか，という疑問に答えを出すためには，その変化が明瞭である基礎的運動要因よりも運動指令が生成されるまでの内的過程に注目しなくてはならないといえる．本論文ではこの内的過程に焦点をあて，学習によりこれがどのような変化を経て技能獲得へと至るのかを論じる．

2.2　内的過程の役割

我々があるタスクのために運動を生成しようとする際には，まずタスク達成のためにどのように身体を動かせばよいかが分かっていること，そして，その通りに身体を動かせること，が必要である．また，目標通りに身体を動かすには自分がどのような動作をしているかという情報が不可欠である．例えば，ボールを的に向かって投げる場合，どのように腕を振って投げるか，が多少なりとも分かっていなければボールを的に当てることはできないし，わかっていてもその通りに動作できなければ同様に的に当てることはできない．このような「行おうとしている運動に関する情報」と「自分が行っている運動の情報」の取り扱いこそが，運動技能の内的過程の役割と考えられる．

認知心理学の分野では，ある動作がなされる以前にその動作に対応する抽象的なプランがすでに用意されていると考えられており，そのプランは運動のプログラム，または行動のプログラムと呼ばれている．運動のプログラムは筋肉運動には直結しないので，実際に運動を行う際は，状況に応じて抽象的なプログラムをより具体的な形に変換し，状況に見合った運動を生成していると考えられている．実際に運動を生成するまで，つまり抽象から具体までの変換には動作に対応した複数の階層が必要となる．

一方スポーツ競技の場では,「行おうとしている運動に関する情報」を “イメージ” という言葉で表現する事が多い．スポーツ選手やそのコーチは「イメージが固まっていない」「イメージ通りの動きができた」というように，イメージという言葉を使って運動技能を評価する．前者は，タスクに対してどのように身体を動かせばいいのかまだはっきりとわかっていない状態，後者は意図通りの運動を実際に行うことができた状態を指す．

ある動作を行う際，その動作に対する表象がすでに準備されている，とする点で運動のプログラムも運動のイメージも共通の考え方である．本論文では，スポーツ競技における例を挙げることが多いため，この準備されている表象の事を運動のイメージと呼ぶように統一する．

さて，“イメージ” という言葉を使うと，常識的には心の中に思い浮かべられる絵のようなもの，あるいは動画みたいなものと考えられてしまいやすいが，ここでいう運動のイメージとは単に目標となる運動を映像的に捉えただけのものではなく，運動感覚に起源を持つ成分を含んだものである [3]．例えば，スポーツ選手がトレーニングの一環として行うイメージトレーニングでは，理想的な運動のイメージを想起することでパフォーマンスの向上を目指す．このとき，選手が想起するイメージは，その運動を第三者的に観る観察イメージと自分が実際にその運動を行っているように想起する行動イメージの 2 種類存在し，その 2 つを必要に応じて切り替えていると言われている [4]．また，佐々木ら [5] は，我々漢字圏の人間が指で字を空中に書くことで，その字を覚える，もしくは思い出す行動

“空書” に着目し，イメージの運動感覚的成分について論じている．

実際に人間がどのような形で運動指令を生成しているかという事はわかっていないが，このように考えると，運動技能の内的過程とは “イメージ” というフォーマット上で目標運動に関する知識を観察イメージ，行動イメージを経て具体的な運動指令へと変換させていく機構であると考えられる．

一方,「自分が行っている運動の情報」についても同様で，これも身体がどのような動作をしているかという体性感覚のみによって形成される情報ではなく，視覚やときには聴覚など様々な感覚入力を統合した上で形成される．この情報を“ 身体イメージ ” や“ ボディイメージ ” と呼ぶ．例えば，我々は今自分がどのような姿勢を取っているかということを第三者的な観点から映像のように想起することが可能である．つまり身体イメージもまた運動感覚だけでなく，視覚的成分を含んでいると言える．

この身体イメージは行動イメージ，もしくは行動イメージから運動指令への変換の修正に利用される．先ほども述べたように，目標がわかっているからといって必ずしもその通りに動けるわけではない．目標と合致する観察イメージを持っていたとしても，それを運動指令に変換するまでの変換が正しくなければ目標通りには動けないからである．このとき,「行おうとしている運動に関する情報」と「自分が行っている運動の情報」という 2 つの情報がそれぞれ単に，視覚的情報と運動感覚的情報というお互いに隔絶した情報ではなく，イメージという共通したフォーマット上にある情報であるため，修正が可能なのだと考えられる．

2.3 学習

我々は学習を通じて様々な運動技能を獲得していく．その際，学習者のイメージは何によってどのように形成され，変化していくのか．ここでは具体的に，一般に挙げられる運動学習の方法である，試行錯誤，見まね，教示の三つについて順に考察する．

2.3.1 試行錯誤による学習

観察イメージ形成の方法として，まず考えられるのは，求める運動の結果から必要な運動を逆算する方法である．例えば，ラケットでテニスボールを特定の方向に打つ運動を考える．我々は球体にどのようにものを当てるとどのような方向に動くか，ということについて経験的にある程度知っている．その知識を基準に飛んでくるテニスボールにラケットを当てる．狙った場所に飛べばそのイメージを記憶する．飛ばなかったら，そのイメージを棄却し，肘の角度や腕を振る方向を変え，ラケットがボールに当たる際の角度を変える．この試行錯誤を繰り返すことにより，運動のイメージが形成される．つまり，実際に

身体を動かすことで，良い結果の出る運動を探し，その時得られる身体イメージから運動のイメージを形成していく手法である．しかし，この方法では，複雑な運動を要するようなタスクの場合に対応できない場合がある．そこで従来の運動学習では，試行錯誤による学習だけではなく，後に述べるような見まねや教示を取り入れて学習の促進を図るのが一般的である．

試行錯誤による学習にはもう一つの側面がある．それは持っている運動のイメージ通りの動きができるようになるための学習である．正しい運動のイメージを持っていたとしても，実際にそれを具体的な運動指令に変換できるとは限らない．言い換えれば，自分が行っていると考えている運動と実際に行っている運動とが一致しているとは限らない．自分が意図していた通りに動作できるようになるためには，具体的な運動指令のパラメータの微調整が必要になる．この微調整を実際に試行錯誤しながら運動することによって行うのである．

2.3.2 見まねによる学習

我々がある運動を初めて行う際には，必ずといっていいほど手本が存在し，その見まねを行うことから学習を始める．また,「上手い人の動きを良く見る」というのは，スポーツ上達の秘訣として種目を問わず広く知られている．このことからも，運動学習には他者の運動の観察が欠かせないことがわかる．

一般に，他者の運動指令を直接観察することはできない．また，仮に何らかの方法で運動指令を観察できたとしても，身体のパラメータが異なるのでその通りの運動を生成できるとは限らない．このことから，学習者が他者の運動の観察から得るのは具体的な運動指令ではなく，より抽象的なもの，つまり運動の観察イメージであると考えるのが自然である．

運動は身体の各部分の動きの組み合わせから生成されるため，これを観察する際，全ての部分を同時に観察することはできない．そして，記憶容量の関係で一度に覚えられる観察イメージにも限りがある．そのため，見まねを行うためには，何度もその運動を観察する必要がある．また，見まねを行う際，他者の運動のどの部分を観察するかは学習者次第なので，観察から得られる運動のイメージには個人差があるはずである．このことから見まねを通じて，技能を獲得するにはその運動の重要な点を効率良く観察するのが上達の早道であると考えられる[6]．

2.3.3 教示による学習

運動に限らず何かを学ぼうとするとき，学習者が単独で学習を行うより，そのタスクに習熟した他者の教示を伴って学習を行う方が効率が良い．しかし，他者の運動の観察の場

合と同様で，習熟者がどのような運動指令を生成しているかを他者に伝えるのは困難であり，仮にそれができたとしても，その苦労に見合うだけの効果が得られるとは考えられない．

学習者が単独で学習を行う場合と比較して，教示者付きで学習を行う場合の有効な点は，適切な運動のイメージの形成を促すことができる点である．教示者は的確な言語教示を通じて，より正しい運動のイメージを学習者に伝えることができる．また，先に述べた見まねの場合においても，教示者が良い手本を見せ，その際，どこに着目するべきかを指示することで効率良く運動のイメージを形成させることができる．しかし，学習者がどのようなイメージを持っているか，は教示者はもちろん学習者自身にもわからないことが多いため，悪いイメージの修正は 0 からイメージを作り出すよりも困難である．（一度身に付いてしまった悪い癖を直すには，0 から良い癖を身に付けるより時間がかかる）

もう一つ教示者付き学習の有効な点は，学習者の運動を客観的に評価できる点である．ここまでに述べたように，ある運動を行うときは，その動作に対応した運動のイメージをもとに運動が生成される．しかし，その運動が学習者の意図した通り，つまり運動のイメージと一致する運動であるとは限らない．そうした場合，学習者が理想と現実の差に気付かなければ学習は進まない．このとき，教示者が学習者の運動を客観的に評価し，この不一致を指摘することができれば，学習を再び進めることができるはずである．しかしながら，学習者がどのようなイメージを持ってその運動を行っているかを教示者が直接知ることはできないため，教示者は学習者の持つイメージを推測しなければならない．総じて言うと，学習者がどのようなイメージを持っているかを正しく推測できる者ほど，優秀な教示者であると考えられる．

2.4 上達の条件

運動学習では，単純に学習に費やした時間と比例して，技能が向上していくわけではない．特に，スポーツ競技の場では，技能は徐々に向上するのではなく，ある時に壁を越えるように急激に向上すると言われている．このことから，技能の向上のためにはいくつかの上達の条件を満たさなければならないと考えられる．この上達の条件がどのようなものか，はここまでに述べた運動指令生成までの過程を考えると分かり易い．

まず，第一の条件として，獲得したい技能の観察イメージを求める技能レベルに必要な程度で明瞭に持っている必要がある．明瞭な観察イメージを得るためには，他者の運動（手本）の観察を繰り返さなければならないし，実際に学習者が身体を動かして，その手本のどの点が重要であるかを知らなければならない．

次に，学習者が持っている観察イメージ通りに運動を生成している必要がある．イメージ通りの運動を生成していない，という事はすなわち，目的とした運動と実際に行ってい

る運動とにズレがあるという事である．当然，目的とした運動は，その技能に必要な運動であるので，目的と実際のズレを無くさなければ技能の向上は見込めない．

この 2 つの条件は，それぞれの変化を表から知る事が困難なので，技能向上のためにどちらの条件が欠けているのかを判断することも同様に困難である．しかしながら，そもそも目標がわかっていないのか，それとも目標はわかっているがそこに辿り着けないのか，が明確でない状態はしばしば学習者の学習効率を著しく下げ，俗に言うスランプに陥ることとなる．

2.5　学習者の分類

運動技能について議論する際，学習者のレベルを判別するために初級者，上級者という具合に学習者を習熟度に合わせていくつかに分類するのが一般的である．ここでは本稿におけるこれらの分類について具体的に述べる．

まず「初心者」「初級者」について述べる．似たようなニュアンスの言葉であり，扱う者によって定義がまちまちの場合もあるが，基本的に「初心者」はその運動に対する経験が全くないもの,「初級者」はある程度の経験があるものを指す．本論文ではこれを受けて,「初心者」をそのタスクに対する経験が全くないもの,「初級者」を経験があり，そのタスクにおける最も基本的な運動を達成することができるもの，とする．

次に「上級者」について述べる．ここでいう「上級者」とは単に初級者に比べて学習回数，学習時間が多いものを指す言葉ではなく，学習を行った環境と異なる未経験のタスクにおいても，事前の知識を利用し対応できるもののことを指す．

スキーを具体例に挙げれば，初心者はスキーを履いたこともないもの，初級者は緩く圧雪された斜面なら自由に滑ることのできるもの，上級者は荒れていたり，急であったりする斜面であっても自由に滑ることのできるもの，という具合に分類する．

この条件は，実際のスポーツの世界において上級者の応用力や外乱に対する安定性が，初級者のそれに比べ高いことからも妥当であると考えられる．また，上級者が行う複雑な運動を達成するために，いくつかの基本的な運動を習得する必要があることを考えると，「ある運動に関する知識・経験の利用」は運動学習において，重要な概念であることがわかる．

2.6　運動イメージの次元

先にも述べたように，他者の運動を観察することは運動学習において重要な要素の一つである．しかし，同じ運動を観察しても，その運動に習熟した者が観察するのと，全く経験のない者が観察するのでは，その見え方は大きく異なる．例えば，スポーツ上達のため

の手段として，学習者が自分の運動の映像と上級者の運動の映像を見比べる事があるが，このとき，学習者の技能が低いと，自分と見本との運動の差を見つける事ができない場合がある．また，上級者とそれよりさらに高い技能を持つ超上級者との間にある運動の差を評価するには，観察者にもその運動に対する高度な理解が必要とされる．このことから筆者は，運動イメージには次元のようなものがあるのではないかと考えた．

先ほどの例をこの考えに沿って考察してみよう．学習者が他者の運動を観察する際，学習者は自らの持つ軸（次元）でその運動を理解しようとする．その結果，その運動の特性（重要点）の一部が学習者に見えない次元に存在する場合，それに気付くことができないのだと考えられる．また，運動の評価を行う場合，上級者と超上級者との運動の差は技能の低い者には見えない高次元に存在するので，その評価を行う者には運動を行う者と同程度の技能が必要とされるのだと考えられる．

これは観察に限ったことではなく，運動を実行する際にも同様である．ある運動を行おうと試行錯誤を繰り返しているとき，学習者は自らの持つ次元からなる探索空間内で目標に合致する運動を探していると考えられるが，この時その運動が探索空間内に存在しなければ，目標を達成することができない．このように，今まで認識していなかった運動の新たな次元に気付くことは高度な技能獲得に不可欠な要素である．

第 3 章

理論モデル

この章では、第 2 章に述べた運動技能についての考察をまとめ、技能獲得過程の理論モデルをダイアグラムとして提示する.

3.1　運動学習ダイアグラム

前章で述べた運動学習についての考察をもとに作成した運動学習ダイアグラムを図 3.1 に示す.

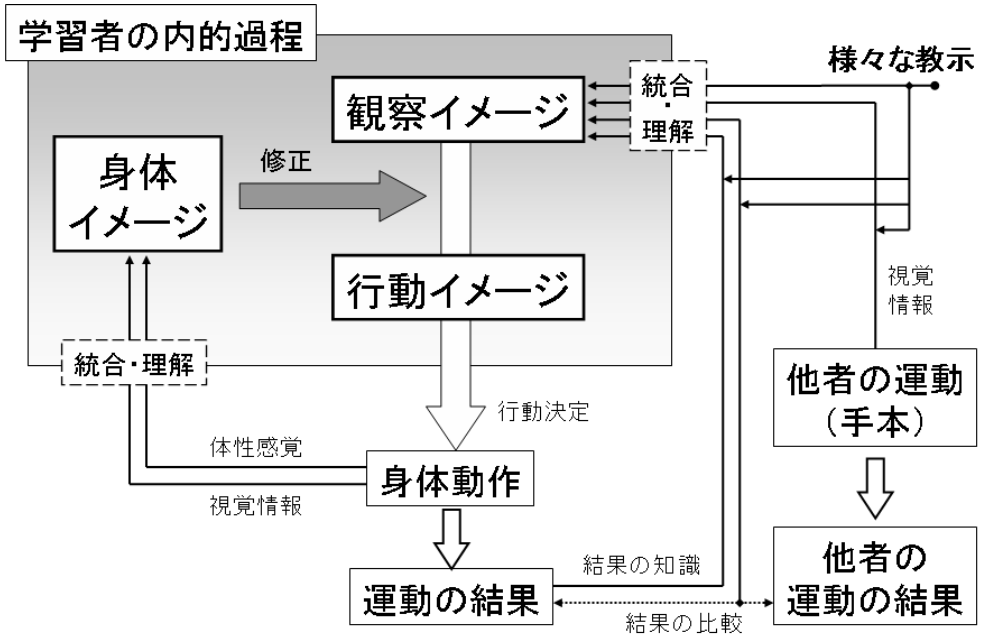


図 3.1　運動学習ダイアグラム

前章でも述べたように運動指令は運動イメージ内の観察イメージから行動イメージを経て生成される．つまり、初めてその運動を行う場合はまず最初に何らかの方法で観察イメージを形成しなければならない．

観察イメージは、他者の運動の観察や教示者からの教示、時には他者からの言葉による説明、によって形成される．このとき、同じ運動を観察した、もしくは同じ教示を受けたからといって、全ての学習者が同じ観察イメージを形成するとは限らないことに注意したい．先にも述べたように、学習者が持つその運動に関する知識や習熟度によって、つまり学習者の持つイメージの次元によって、このとき形成されるイメージは異なる．また、単純に他者の運動を観察する際、学習者がどこに注意を注ぐかによっても形成されるイメージは異なる．

次に、観察イメージは行動イメージを経て運動指令となり、学習者は実際にその運動を行う．同時に、学習者は運動を行う際の体性感覚や視覚情報などから身体イメージを形成する．(なお、形成された身体イメージが実際の運動を正確に反映しているとは限らない.) 学習者は形成された身体イメージから自分が実際にどのような運動を行ったかを知り、イメージ内での変換の機構を修正する．

最後に、運動の結果をもとに観察イメージを修正する．まず学習者が自分が行った運動の結果を評価する．つまり、自らが持つ観察イメージがどのような結果を残したかを検討する．その結果、今の観察イメージに不足がある場合、観察イメージを修正する．この際、例えば他者の運動の結果と比較を行うなどして、検討を促進させることも考えられる．

以上が学習者が単独で運動学習を行う際の手順である．これに教示者が加わる場合、その教示によって、運動の結果の評価がより正確になる、他者の運動を観察する際どの部分を観察するべきかの誘導がある、などが期待され、その結果観察イメージ修正の精度が向上すると考えられる．

3.1.1　運動学習のサイクル

ここではスキーを例として，全く経験のない初心者が初級者になるまでの過程を前述のダイアグラムに沿って考察したい．ここで，スキーにおける初級者とは，緩やかな斜面を真っ直ぐスピードを調節しながら滑り降りることのできる者とする．

初心者が「緩やかな斜面を真っ直ぐスピードを調節しながら滑り降りる」ためにはプルークファーレンという滑り方を身に付けるのが早道となる．プルークファーレンとは，板を“ハ”の字の形にして（なお，これをプルーク姿勢という）雪面からの抵抗を増し，スピードを調節する技術である．

プルークファーレンには着目するべきいくつかのポイントがあるが，この姿勢で最も特徴的なのは，やはりハの字にした板である．よって学習者の学習最初期には，他者の観察

から「板をハの字にして滑る」という運動のイメージがまず形成される．

次にこの「板をハの字にして滑る」というのを目標に実際に運動を行う．このとき，学習者の中には板をハの字に開く事のできない者もいる．これは目標通りの姿勢になるために，どのように身体を動かしたらいいかがわからない状態，つまり，運動のイメージから具体的な運動のプランへの変換がうまくいかなかったのだと考えられる．こうした状態から脱するために，学習者は体性感覚や視覚から自分の行った運動についての情報を得て，運動のイメージからプランへの変換の機構を修正しているのだと考えられる．

最後に，学習者による運動の結果の評価が行われる．プルークファーレンの場合，“ハ”の字の大きさによってスピードの調整を行う．例えば，ハの字に足を開く事ができても，求めていた以上にスピードが出てしまった場合，より大きく足を開くように目標を変更，つまり運動のイメージを修正する．また，運動のイメージの修正は前述したような他者の運動の観察や教示者からの運動の評価によっても修正される．

学習者はこうした作業を繰り返すことで,「緩やかな斜面を真っ直ぐスピードを調節しながら滑り降りる」ために必要な技術の運動イメージを獲得し，そのイメージ通りに運動を生成できるようになると考えられる．

3.1.2　運動技能の差

前節では，スキーを例にとって初心者が学習を経て，初級者になるまでの過程について考察した．ここでは引き続きスキー，プルークファーレンを例にとって初級者と上級者の技能の差についての考察をダイアグラム上で行う．

初級者と上級者，両者の技能には当然ながら歴然とした差が存在する．しかし,「緩やかな斜面を真っ直ぐスピードを調節しながら滑り降りる」というタスクに関しては，初級者も上級者も「達成できる」という点では違いはなく，両者の技能差は結果には表出しない．

両者の技能差が結果に表れるのは，タスクの難度が上がったときである．例えば，前述のタスクの斜面設定を緩やかな斜面から急な斜面に変更したとすると，同じプルークファーレンを行ったとしても，初級者はタスクを達成することができない．なぜ，同じ滑り方をしているのに，上級者には出来て初級者には出来ないのか．それはプルークファーレンという滑り方に対する理解度の差，つまり運動イメージの差に原因があると考えられる．

前節で，初心者が初級者になるまでの過程について考えた際，学習者は他者の動きの観察から「板をハの字にする」というイメージを得て，その通りに行動できるように練習を行った．そして，その結果タスクを達成できるようになった．しかし，前にも触れたようにプルークファーレンにはこれ以外にも，いくつかの重要な点が存在する．(例えば，斜面に対して垂直方向に体軸を取ると板により力が伝わりやすくなり，結果急な斜面でも十分

に速度を調節できるようになる.）上級者は初級者の獲得したもの以外にもこうした重要点を獲得している．つまり，初級者に比べてプルークファーレンという滑り方に対して，より細かくより明瞭な形で運動のイメージを持っており，またそのイメージ通りの運動を生成できていると考えられる．

初心者が獲得した運動のイメージでは，緩やかな斜面という比較的粗い運動でも許容される環境には対応できても，急な斜面のように精密な運動を要する環境には対応できないのである．

3.1.3　上級者になるためには

前節では初級者と上級者の技能の差についての述べた．ここでは初級者がいかにして上級者になるかについて引き続きスキーを題材に考察する．

初級者と上級者の技能の差がタスクの難度の高い環境で表出することは前にも述べた．このとき，初級者が上級者の持つ高い技能を身に付けるにはどうしたら良いのか．もっとも単純に考えられるのは，最初から急な斜面で学習することである．しかし，この方法では高い技能を身に付けることができないどころか初級者としての技能すら身に付けられない可能性がある．なぜなら，急な斜面を滑るための技能には，緩やかな斜面を滑る技能に比べて押さえなければならない重要点が数多く存在するからである．さらに，そのうちのいくつかは以前にも述べた運動のイメージの次元で言えば，初心者には見えない次元に存在しており，これら全てについての観察イメージを獲得する，つまり必要な多数の重要点を押さえることは困難である．また，実際に運動を行う際にも探索空間が広大になってしまい，タスク達成に必要な運動を生成するのが極めて難しい．そのため，一般的に初心者は探索空間の狭い簡易な環境で学習を開始するのである．

次に考えられるのは単純に学習時間を長く取ることである．しかし，この方法でも高い技能を獲得するのは不可能であるといえる．初級者は緩やかな斜面では「スピードを調節して滑る」というタスクを既に達成できている．よって，単純に学習を進めても観察イメージの修正が行われることはない．つまり，初心者は緩やかな斜面において一種の局所解に陥っていると考えられる．

この状況で学習を促進させるためには，学習者の “気付き” が必要となる．つまり，今の解，今の技能に満足せず，足りない部分を探すというモチベーションが無ければ，局所解から抜け出ることはできない．こうした “気付き” は実際には，教示者からの指摘や実際に急な斜面を滑って失敗することで促される．

こうして自らの技能の穴に気付いた学習者は以前とは別のモチベーションで，緩やかな斜面を滑ることになる．つまり，今までは「真っ直ぐスピードを調節しながら滑り降りる」という結果だけを求めており，それに必要な精度だけでプルークファーレンを取得し

ていたが，これからは緩やかな斜面に必要な精度以上のプルークファーレンのイメージを身に付けようとして学習を行うのである．この場合，今まで初級者の中ではただ足をハの字にして滑るだけだったプルークファーレンを板の真上に乗るように，荷重が板の中心にかかるように，といったように練習することになる．

“気付き” は学習者単独でも起こるが，その後どのような目的で学習を行えば良いかを学習者単独で探すのは困難である．そうした側面からも学習者単独の学習よりもその運動に習熟した教示者付きの学習の方が効率が良いということが言える．

3.2 運動学習の理論モデル

前節までに運動学習についての一般論を述べ，それをまとめた運動学習ダイアグラムを提案した．第 4 章では，このダイアグラムをシミュレーションモデルとして実装し実験を行うことでダイアグラムの妥当性を検討する．しかし，前述のダイアグラムはシミュレーションモデルとして実装するには抽象的である．そこで，ここでは運動学習ダイアグラムをフィードバックという概念を用いて簡略化したモデルを提案する．

今までに述べたように我々が運動学習を進める過程では，運動のイメージや運動のイメージを具体的な運動指令に生成する機構を修正していく．このときに運動結果から得られる情報のことをフィードバックという．フィードバックとは，ある運動において，目標値とパフォーマンスとの間の差異についての情報であり，学習者はこの情報を利用して学習を進める．このフィードバックには内的フィードバック，外的フィードバックという 2 つの種類が存在するので，以下でそれぞれの特徴について述べる．

特殊な装置や方法無しに学習者が直接知覚できるフィードバックのことを内的フィードバックと呼ぶ．我々がある動作を行ったとき，体性感覚や自分の四肢の動きを視覚で捉えることで，自分がどのような動作を行ったかを知ることができる．この運動の結果についての情報が内的フィードバックである．

基本的に体性感覚と視覚情報のみで，自分がどのような運動を行っているかを完璧に知るのは困難なので，結果内的フィードバックも不正確なものとなる．また，内的フィードバックの正確性はその運動の習熟度に比例すると言われている．

学習者が直接知覚できる内的フィードバックに対して，運動結果からの外的情報のことを外的フィードバックと呼ぶ．外的フィードバックはさらに結果の知識，パフォーマンスの知識の 2 つに分類することができる．

例えば，的に向かってボールを投げたとき，そのボールが的に当たったかどうか，など運動の目標に対する行為の結果について得られた情報のことを結果の知識という．結果の知識は，その結果をどのように評価するかによって得られる情報が異なる．例えば，投げたボールが的から外れた場合,「外れた」とだけ認識するか,「どちらの方向にどれだけ外れ

た」と認識するか，でその後の学習に及ぼす影響は大きく異なる．

学習者が行った運動について，教示者がアドバイスをするように付加的に与える情報のことをパフォーマンスの知識という．また，教示者によるものだけでなくても，例えば学習者が行った運動を撮影したものを，学習者が改めて客観的に観察することでもパフォーマンスの知識を得ることができる．

以上をまとめた簡略化した運動学習についてのモデルを図 3.2 に示す．

身体イメージを形成するために知覚される体性感覚，視覚情報は学習者が直接知覚できる内的フィードバックである．一方で，運動の結果から得られる学習者による結果の評価や他者の運動の観察によって得られるその運動に関する視覚情報や教示者による教示は全て外的フィードバックにあたる．

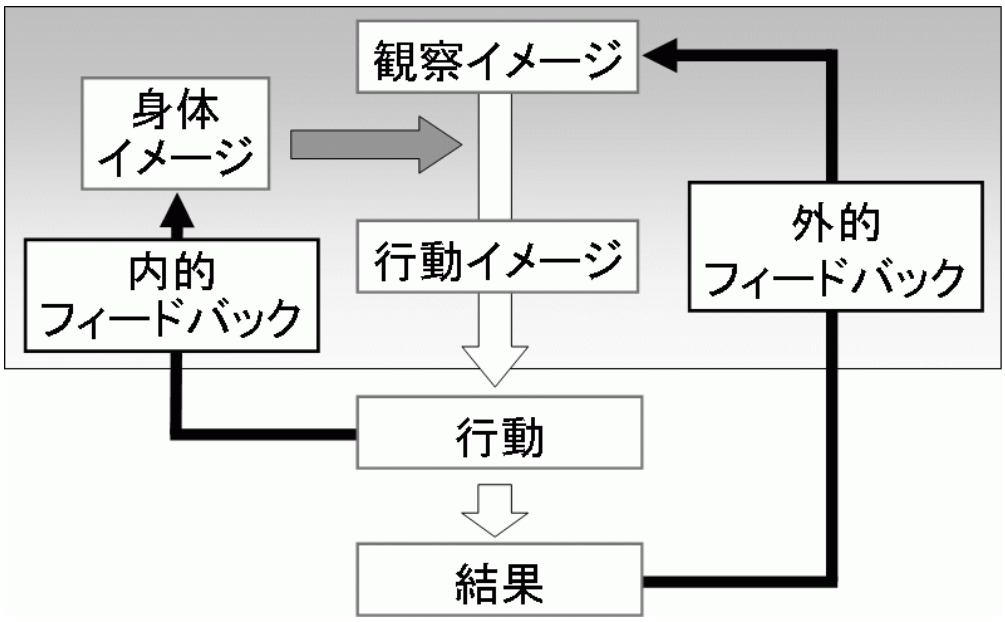


図 3.2　運動学習の理論モデル

第 4 章

シミュレーションモデル

ここでは，前節の理論モデルをシミュレーションモデルとして実装したものを提示し，それらの実装について解説する．

4.1 シミュレーションモデルの概要

前章で提示したダイアグラムをもとに，運動結果に基づく報酬による強化学習ベースのシミュレーションモデルを作成する．モデルの概要を図 4.1 に示す．

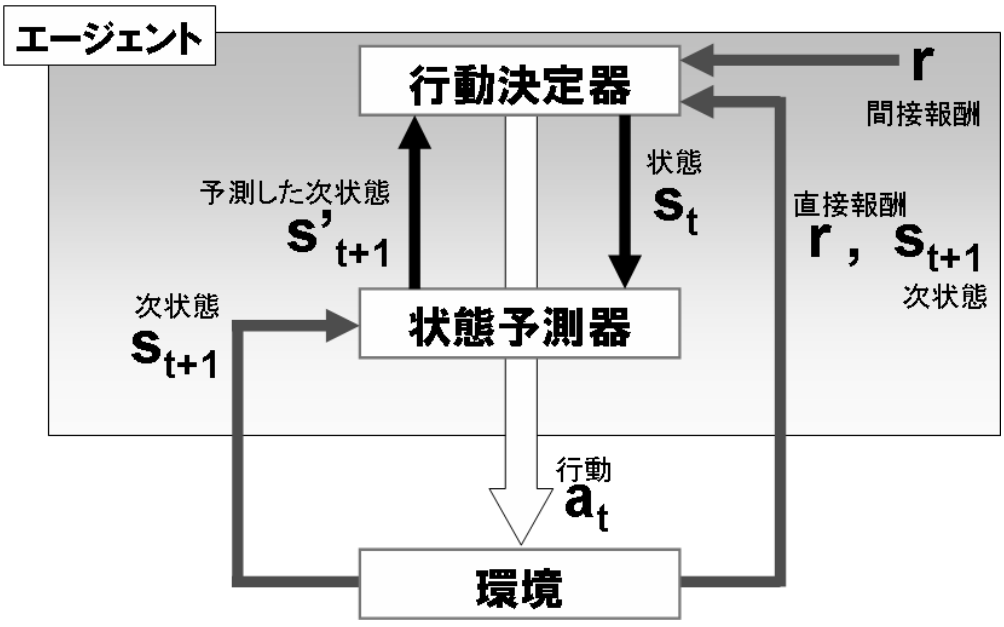


図 4.1 シミュレーションモデル

このモデルは，状態予測器と行動決定器の 2 つの学習器からなる．状態予測器はその環境の系を近似する内部モデルを持ち，状態 s_t にある行動を取った後の状態 s_{t+1} を予測する．行動決定器はそのタスクにおいて，どのような状態にあるのが良いかという指標を状態価値関数として持っており，状態予測器から受け取った予測結果 s'_{t+1} から最も良い次状態となる行動 a_t を決定する．

行動後，環境から状態予測器へ実際の次状態である s_{t+1} が，行動決定器へ s_{t+1} と報酬 r が返される．状態予測器に返される実際の次状態 s_{t+1} は内的フィードバック，行動決定器へ返される報酬 r は外的フィードバック，にそれぞれ対応する．

状態予測器は自分が予測した次状態 s'_{t+1} と環境から返された実際の次状態 s_{t+1} との誤差を用いて学習を行い，予測の精度を高める．一方，行動決定器は環境から返された報酬 r を用いて学習を行う．報酬 r は状態 s_t のとき取った行動 a_t に関する評価を表し，行動決定器の持つ状態価値関数を変化させる．

4.2 行動決定器

行動決定器では，理論モデルにおける観察イメージの形成と修正の役割を実装している．以下では行動決定器の実装について具体的に述べる．

4.2.1 観察イメージと行動決定器

前節で述べたように，観察イメージの形成および修正には運動の結果に対する外的フィードバックが用いられる．このとき，外的フィードバックは明確な正解を示しているわけではなく,「こうした方が良い」「ああした方が良い」といった結果の評価の役割を果たしている．これは，外部から与えられる報酬や罰などの強化信号をもとに学習を行う，という強化学習の考え方に良く似ている．報酬は教師あり学習の正解とは異なるが，目的に合致した行動の結果として与えられる．つまり,「どのような状態でどの程度の報酬を与えるか」という報酬のデザインによって学習エージェントの振舞いは大きく異なることになる．

ある運動を行ったとき，結果としてはそのタスクに失敗してしまっても，その一連の運動の中に良い運動があったかもしれない．このようにその運動を結果だけ見て「悪い」と判断するか，途中の過程を考えて「良い」とするかは学習者によって異なる．また，目標となる運動を他者が行っているのを観察する際も，どの点に着目するかによって得られる運動の評価は異なるし，学習を行う際，教示者によるアドバイスがあるかないかによっても異なる．

本稿で提案するモデルでは，タスクの達成そのものを表す報酬に加えて，タスクに関連

する別の仮想的な報酬をエージェントに与えることで，こうした運動の評価の差異を実現する．以下では，タスクの成功，失敗そのものを表す報酬を「直接報酬」，タスクに関連する仮想的な目標の達成度を表す報酬を「間接報酬」と呼ぶ．

直接報酬を「タスクの達成条件を記述する報酬」とすると，間接報酬はタスクを達成するためには直接関わらない，タスクをいかに上手く解くかに関わる「コツ」のようなものと言える．

4.2.2 状態価値関数を用いた行動決定器の実装

強化学習では，報酬を評価してそれを最大化することで学習を行う．現在の状態がどのくらいよいかを計る関数として，価値関数を考える．“どのくらいよいか” は，将来にわたって得られる報酬により定義する．すなわち，状態価値関数は現在の状態の評価の指標である．状態 s の状態価値関数 $V(s)$ は，状態 s から全てのステップにおいて行動したときに期待できる報酬の和として，以下のように定式化できる．

$$V(s) = \sum_{k=0}^{\infty} \gamma^k r_{t+k+1} | s_t = s$$

先に述べたように，このモデルにおける報酬は運動の評価に対応している．学習者（エージェント）は，理想の運動の観察イメージを直接知ることができないので，運動の評価によって現在持っている観察イメージを徐々に修正して，理想のものに近付けていく．そこで本研究では，最も一般的な強化学習のアルゴリズムである TD 学習を用いる．具体的に，状態価値関数は以下のルールで更新する．

$$\begin{aligned} V(s_t) &\leftarrow V(s_t) + \alpha \left[r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \right] \\ &\leftarrow V(s_t) + \alpha \delta_t \end{aligned}$$

α は学習率，γ は割引率，δ_t を TD 誤差である．TD 誤差は，理想の価値関数と現在の価値関数との誤差，すなわち外的フィードバックを考慮した理想の観察イメージと現在の観察イメージの差を表している．

なお，本稿で提案するモデルでは，ガウス型動径基底関数を使って状態価値関数 $V(s)$ を以下のように連続的に記述する．

$$\begin{aligned} V(\boldsymbol{s_t}) &= \sum D_i f(\boldsymbol{s_t}, \boldsymbol{\mu}) \\ &= \sum D_i \exp\left(-\frac{1}{2} cov^{-1} [\boldsymbol{s_t} - \boldsymbol{\mu}] \right) \end{aligned}$$

ここで，cov はガウス型動径関数の広がりを表す行列，μ は動径関数の中心位置を表すベクトルである．

このように状態価値関数を記述する場合，状態価値関数の更新は，$V(s_t)$ を直接更新するのではなく，上式中の係数 D_i を以下の更新則で更新することになる．

$$\begin{aligned} D_i &\leftarrow D_i + \alpha \left[r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \right] f(\boldsymbol{s_t}, \boldsymbol{\mu}) \\ &\leftarrow D_i + \alpha \delta_t f(\boldsymbol{s_t}, \boldsymbol{\mu}) \end{aligned}$$

4.3 状態予測器

状態予測器では，試行錯誤を繰り返し運動のイメージと実際の運動との差を縮めていく機構を実装している．実際の運動では，自分の運動の結果を体性感覚や視覚情報を通じて得るため，内的フィードバック，理想と実際の差は正確なものとならない．しかし，本稿で提案するシミュレーションモデルでは実際に自分が行っている運動は正確にわかるものとする．

4.3.1 内部モデルと状態予測器

あるタスクを達成するためにどのような運動をするべきかがわかっていても，実際にその目標通りの運動を生成できなければ，当然そのタスクを達成することはできない．また，目標通りの運動を生成するためには，現在の自分の状態からどのような動作を行えばその運動ができるか，という事が予想できなければならない．学習者は試行錯誤を通じて得た実際の運動と予想の誤差を用いて，この予想のための内部モデルを形成，修正しながら，実際の運動を目標の運動に近付けているのだと考えられる．

前にも述べたように，初級者が学習した環境でしかタスクを達成できないのに対して，上級者はどのような環境でも事前の知識を活かして対応することができる．これは，初級者に比べて上級者の持つ内部モデルがより環境のパラメータ（例えば，ボールの重さやゴールまでの距離など）に依存しない普遍的な形，つまり，より環境の構造に近い形を成しているからだと考えられる．

4.3.2 ボトムアップ型学習アルゴリズムを用いた状態予測器の実装

本研究では，学習者が試行錯誤を通じて内部モデルを形成していく機構をボトムアップ型学習アルゴリズムを利用して実装した．

ボトムアップ型学習アルゴリズム（以下 BU 法）とは，下位層から順次特徴素子を生成してネットワークを自己組織化させながら，目的の関数を実現しようとするものである [7]．すなわち，与えられた入出力関係を模擬できるようになるまで，ネットワーク構造を積み上げていくという形をとる．

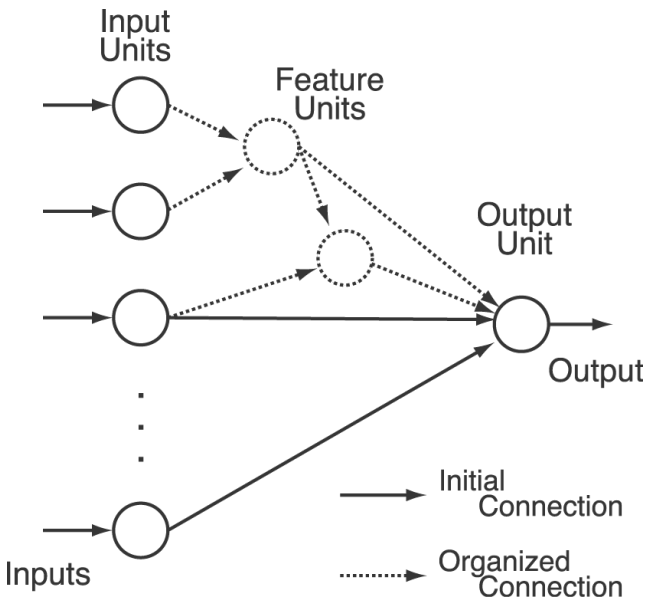


図 4.2　ボトムアップ型学習アルゴリズムの概念図

ネットワークを構成する素子は，入力素子，特徴素子，出力素子の 3 種に分けられる．入力素子は外部からの入力 (この場合は状態変数) をそのまま出力する．特徴素子は結線された他の素子からそれぞれ信号を受け取りその積を，出力素子は結線されたあらゆる素子からの線形和を出力する．

学習初期では入力素子のみを利用し，素子間の結合係数を修正して，出力誤差を小さくしていくように学習を行う．学習を一定回数繰り返しても出力の平均誤差が一定量以上減少しない場合，入力素子，特徴素子の中から 2 つの素子を機械的に選び，その積を取る形で新たな特徴素子を作成し，再び学習を行う．また，生成した特徴素子がすべて有意義であるわけではないので，情報処理の負荷を小さくするために素子数が一定量を超えた場合，最も結合係数の低い素子を不要と考え，その素子から出力素子への結合を切断する．

4.4　シミュレーションモデル上にみる技能の差

ここまでに説明した通り，エージェントは行動決定器，状態予測器の学習を行いながら，この 2 つの学習器を使って，タスク達成を目指す．つまり，この 2 つの学習器の性能がそのエージェントの技能であるといえる．前章のダイアグラムを使って述べたように，初級者と上級者の技能の差は，観察イメージの差とその観察イメージ通りの動きを生成できるかどうかに起因している．

状態予測器は予測誤差を用いて自律的に学習を行う．一方，行動決定器は外部から与えられる報酬によって学習を行う．つまり，タスクを達成できるかどうかは，タスクの達成条件を正確に反映した報酬が得られるかどうかに強く依存している．単純なタスク（ス

キーの例では緩斜面をまっすぐすべること）であれば，適切な報酬（身体が倒れないとよい評価が与えられる）が与えられれば，タスクは必ず達成することができる．しかし，このような状況で，問題の性質が少しでも変化する（斜面の角度が急になる）と，タスクを達成することは難しくなる．これは，単純なタスクを実現する際に獲得した状態予測器や行動選択器は，そのタスクを実現するのに必要な最低限の能力しかもっていないためと考えられる．

逆に，上級者は，問題の内容が少々変化しても対応できるような状態予測器と行動選択器を獲得していると考えられる．これは，個別のタスクにおいて高い成績を生むような内部モデルではなく，問題の情報構造を反映し問題のバリエーションに対して対応できるような表現構造をもった内部モデルを獲得しているためであると考えられる．すなわち，初級者と上級者の違いの一つに，環境における因果関係を表す内部モデルの汎化能力の違いがあると考えられる．

このような解釈の下では，高い技能の獲得には汎化能力の高い内部モデルの獲得を促すような仕組みが有効である．本稿で提案するモデルにおいてこの仕組みに相当するものが間接報酬である．

先に述べたように，間接報酬は，学習者自らが設けた，あるいは熟練者に教示によって定められた仮想的な報酬である．間接報酬を受け取った行動決定器は，この報酬が高くなるように行動決定を行なうことになるが，その結果として，学習エージェントは技能獲得に必要な行動を積極的にとるようになる．すなわち，直接報酬のみの下では選択しにくい行動を間接報酬の助けによって選択しやすくするような作用が生じる．

このような状況の下で学習を行なうことにより，獲得される状態予測器も変化することになる．したがって,「適切な」間接報酬を与えることにより，状態予測器が局所的な最適解（つまり，環境構造を反映していない解）に陥ることを防ぐことができれば，その結果学習エージェントはより汎化能力の高い内部モデルを獲得できるようになると考えられる．

第 5 章

実験

ここでは，前章で提案したシミュレーションモデルに倒立振子の制御を学習させ，モデルの妥当性を検討する．

5.1 実験概要

第 4 章で述べたように，学習者の技能はエージェントの持つ状態予測器と行動決定器の性能に対応しており，学習の際与える報酬によってエージェントの獲得する技能は異なったものとなる．

ここでは，直接報酬のみで学習を行うものと直接報酬と間接報酬の 2 つで学習を行うものの 2 種類のエージェントに倒立振子の制御を学習させ，両者の技能の差を検討する．なお，間接報酬については何が適切な値であるかが不明なため，30 種類の間接報酬を用意して総当りに試行する．つまり，計 31 個のエージェントに学習を行わせることとした．

行った実験は 3 つである．まず実験 1 では，単純にそれぞれのエージェントがタスクを達成できるかどうかを見る．次に，実験 2 では学習後の各エージェントの持つ状態予測器の性能を見る．そして最後に，実験 3 では学習後の技能を用いて，より難度の高い他の環境へ適応できるかを見る．

5.2 倒立振子

倒立振子とは図 5.1 に示すような振子を逆さにしたものであり，振子の末端は台車に接続されている．逆さになり不安定になった振子を台車を左右に動かすことで安定に保つように制御するのがこのタスクの目的である．

なお，倒立振子はフィードバック制御の最も基本的な実験の一つであり，また強化学習のデモにも広く使われている．

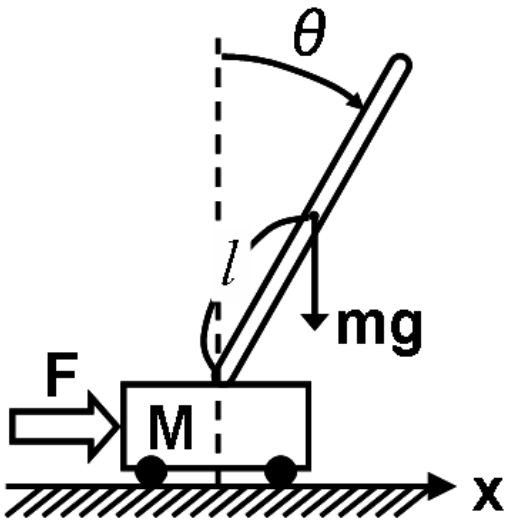


図 5.1　倒立振子

5.2.1　倒立振子の運動方程式

実験で扱う倒立振子の状態 s は 4 つの状態変数 $x, \dot{x}, \theta, \dot{\theta}$ で記述され，これらの変数は以下の運動方程式をもとに更新される．

$$\ddot{x} = \frac{F + ml\dot{\theta}^2 \sin\theta}{m+M} - \frac{ml\cos\theta}{m+M}\left(\frac{g\sin\theta - \frac{\cos\theta}{m+M}\left(F + ml\dot{\theta}^2\sin\theta\right)}{l\left(\frac{4}{3} - \frac{m\cos^2\theta}{m+M}\right)}\right)$$

$$\ddot{\theta} = \frac{g\sin\theta - \frac{\cos\theta}{m+M}\left(F + ml\dot{\theta}^2\sin\theta\right)}{l\left(\frac{4}{3} - \frac{m\cos^2\theta}{m+M}\right)}$$

式中の m, M, l, g はそれぞれ，振子の重さ，台車の重さ，振子の長さ，重力加速度を表す．このタスクではエージェントの取りうる行動は台車を左右のどちらかに押す，という二通りの行動だけで，このとき台車にかかる力の大きさ F は 10 or -10 である．

状態予測器は上式を近似するような内部モデルを持ち，次状態についての予測を行い，その予測情報をもとに行動決定器は価値関数の高い次状態となる行動を選択する．

5.2.2 制御失敗

このタスクの目的は倒立した振子を倒さず安定に保つことである．そのため振子の角度 θ が -1.57 より小さく，もしくは 1.57 より大きくなったとき，失敗とする．さらに，実際には台車を動かすことのできる範囲にも制限があると考えられるので，ここではその範囲を $-2.8 < x < 2.8$ とし，この範囲を超えた場合にも失敗とする．また，振子の速度があまりに速過ぎるとき，台車をどのように動かしてもその状態から安定状態に戻ることはできない．そこで振子の角速度の絶対値が 3.0 を超えたときも失敗とする．

5.3 実験条件

5.3.1 報酬

直接報酬として制御に失敗したときに-100 の報酬を与える．

間接報酬は表 5.1 のように設定した．なお，予備実験によって倒立振子は $x, \dot{x}$ に注意して制御を行うよりも $\theta, \dot{\theta}$ に注意して制御を行う方が制御が上手くいくことがわかっているので，間接報酬は $\theta, \dot{\theta}$ に関するものだけとした．

5.3.2 入力素子

倒立振子の運動方程式は上述のように，$\ddot{x}$ の式と $\ddot{\theta}$ の式の 2 つからなるので，状態予測器の持つ内部モデルも $\ddot{x}$ を予測するものと $\ddot{\theta}$ を予測するものの 2 つとなる．

本来，状態予測器は 4 つの状態変数 $x, \dot{x}, \theta, \dot{\theta}$ とエージェントの行動の結果である F を入力素子として持っており，そこから次状態の予測値を計算するのが理想であるが，倒立振子の運動方程式は sin など周期関数を多数含んでいる．このような周期関数を状態変数の積の組み合わせで近似するのは極めて困難である．また，方程式には $x, \dot{x}$ を含む項は無い．そこで，ここでは便宜的にあらかじめエージェントが持っている入力素子は $\sin\theta, \cos\theta, \dot{\theta}, F$ の 4 つとする．

学習開始時にはこの 4 つの入力素子のみで近似を行う．2000 ステップの学習が終わったとき，平均誤差が 0.02 以上減少しない場合，新たな特徴素子を生成・追加する．なお，出力素子に結線できる素子の最大数は 8 で，特徴素子生成時に素子がすでに 8 個存在する場合，8 個の素子の中で最も結合係数が低い素子の結線を切断し新たな素子を出力素子に繋ぐ．なお，入力素子，特徴素子の結合係数の初期値は $\pm 1.0 \times 10^{-5}$〜2.0×10^{-5} の範囲でランダムに与えた．

ところで，入力素子の積の組み合わせでは，運動方程式の分母に含まれる $\cos^2\theta$ を表す

ことができない．つまり，どんな特徴素子を作ってどれだけ学習しても内部モデルと運動方程式が同様のものになることはない．

表 5.1　間接報酬パターン

間接報酬パターン	報酬内容
0	$\|\theta\| < 0.05$ のとき＋ 10
1	$\|\theta\| < 0.1$ のとき＋ 10
2	$\|\theta\| < 0.2$ のとき＋ 10
3	$\|\theta\| < 0.3$ のとき＋ 10
4	$\|\theta\| < 0.4$ のとき＋ 10
5	$\|\theta\| < 0.5$ のとき＋ 10
6	$\|\theta\| < 0.6$ のとき＋ 10
7	$\|\dot{\theta}\| < 1.0$ のとき＋ 10
8	$\|\dot{\theta}\| < 1.5$ のとき＋ 10
9	$\|\dot{\theta}\| < 2.0$ のとき＋ 10
10	$\|\dot{\theta}\| < 1.75$ のとき＋ 10
11	$\|\dot{\theta}\| < 1.25$ のとき＋ 10
12	$\|\dot{\theta}\| < 2.0$ のとき+5 $\|\theta\| < 0.3$ のとき+5
13	$\|\dot{\theta}\| < 1.75$ のとき+5 $\|\theta\| < 0.3$ のとき+5
14	$\|\dot{\theta}\| < 1.25$ のとき+5 $\|\theta\| < 0.3$ のとき+5
15	$\|\dot{\theta}\| < 2.0$ のとき+10 $\|\theta\| < 0.3$ のとき+10
16	$\|\dot{\theta}\| < 1.75$ のとき+10 $\|\theta\| < 0.3$ のとき+10
17	$\|\dot{\theta}\| < 1.25$ のとき+10 $\|\theta\| < 0.3$ のとき+10
18	$\|\dot{\theta}\| < 2.0$ のとき+5 $\|\theta\| < 0.2$ のとき+5
19	$\|\dot{\theta}\| < 1.75$ のとき+5 $\|\theta\| < 0.2$ のとき+5
20	$\|\dot{\theta}\| < 1.25$ のとき+5 $\|\theta\| < 0.2$ のとき+5
21	$\|\dot{\theta}\| < 2.0$ のとき+10 $\|\theta\| < 0.2$ のとき+10
22	$\|\dot{\theta}\| < 1.75$ のとき+10 $\|\theta\| < 0.2$ のとき+10
23	$\|\dot{\theta}\| < 1.25$ のとき+10 $\|\theta\| < 0.2$ のとき+10
24	$\|\dot{\theta}\| < 2.0$ のとき+5 $\|\theta\| < 0.1$ のとき+5
25	$\|\dot{\theta}\| < 1.75$ のとき+5 $\|\theta\| < 0.1$ のとき+5
26	$\|\dot{\theta}\| < 1.25$ のとき+5 $\|\theta\| < 0.1$ のとき+5
27	$\|\dot{\theta}\| < 2.0$ のとき+10 $\|\theta\| < 0.1$ のとき+10
28	$\|\dot{\theta}\| < 1.75$ のとき+10 $\|\theta\| < 0.1$ のとき+10
29	$\|\dot{\theta}\| < 1.25$ のとき+10 $\|\theta\| < 0.1$ のとき+10

5.3.3 各種パラメータ

環境のパラメータ

- 振子の質量 $m = 0.15$ [kg]
- 振子の重心までの距離 $l = 1.5$ [m]
- 台車の質量 $M = 1.0$ [kg]
- 重力加速度 $g = 9.8$ [m/s^2]
- 台車を押す力 $F = 10.0$ [N]

メタパラメータ

- 行動決定器の学習率 $\alpha_V = 0.1$
- 状態予測器の学習率 $\alpha_F = 0.0001$
- 状態予測器の割引率 $\gamma = 0.95$

状態価値関数のための動径基底関数はガウス型動径基底関数を使用し，x 次元に 6 個，$\dot{x}$ 次元に 8 個，θ 次元に 6 個，$\dot{\theta}$ 次元に 8 個の計 2304 個を $|x| < 3.0, |\dot{x}| < 6.0, |\theta| < 1.7, |\dot{\theta}| < 6.0$ の範囲で等間隔に配置した．また，分散値は基底関数を配置した間隔の 1.5 倍とした．

5.4 実験１：技能の獲得

まず初めに，以下に示すような条件で倒立振子の制御を行わせ，直接報酬のみのエージェント，間接報酬ありのエージェント双方に技能を獲得させる．ここで言う技能とは，学習モデル上の内部モデルと状態価値関数を指す.

学習ステップ条件

$x = 0, \dot{x} = 0, \theta = 0, \dot{\theta} = 0$ を初期条件として，この状態から制御を開始する．制御に失敗すると１エピソード終了とする．50000 ステップの間，制御に失敗しなければ制御成功と見なし，この時点で学習を打ち切る．また，3000 エピソードが経過しても 50000 ステップを越えない場合，制御不可として同様に学習を打ち切る．

直接報酬のみのエージェントと 30 個の間接報酬ありのエージェントのうち 20 個が 50000 ステップの制御に成功した．このことから，間接報酬は単純に付加すれば良いわけではなく，直接報酬と同様に適切な値を選んで付加しなければならないものだということがわかる．

結果

各エージェントの学習の結果を表 5.2 に示す．表中の数字は，そのエージェントが 50000 ステップの制御に成功するまでに要したエピソード数であり，制御不可はそのエージェントが 3000 エピソードが経過しても 50000 ステップの制御ができなかったことを表す．

表 5.2　学習結果

直接報酬のみ	学習結果
	18

間接報酬パターン	学習結果
0	制御不可
1	制御不可
2	324
3	749
4	制御不可
5	2950
6	43
7	718
8	887
9	2764

間接報酬パターン	学習結果
10	458
11	144
12	1396
13	制御不可
14	2401
15	974
16	975
17	758
18	制御不可
19	2974

間接報酬パターン	学習結果
20	1501
21	2788
22	2927
23	制御不可
24	671
25	制御不可
26	制御不可
27	2510
28	制御不可
29	制御不可

5.5　実験２：状態予測器の性能比較

実験１の結果，直接報酬のみで学習を行うエージェントと直接報酬と間接報酬の両方で学習を行うエージェントのうち合計 20 個が 50000 ステップの制御に成功した．これらのエージェントは「制御に成功した」という結果は同じであるが，用いた報酬が異なるので，学習の結果形成される状態価値関数と内部モデルには違いがある．つまり獲得した技能は異なると考えられる．そこで，以下では各エージェントの技能の比較を行う．

ここでは，まずエージェントの獲得した技能のうち，状態予測器の持つ内部モデルの性能，つまり，状態予測の精度を比較する．

実際に倒立振子を制御させて技能の差を評価しようとする場合，その要因が行動決定器にあるのか状態予測器にあるのかが判然としない．そこで，実際に倒立振子の制御を行うのではなく，実験 1 での学習後の内部モデルについて，結合係数を初期化しランダムに状態変数を入力して状態予測器の学習を行わせ，その誤差収束の様子を観察することで，実験１で獲得された内部モデル単体の性能の比較を行った．なお，ランダムで入力される状態変数は以下のように，実際の倒立振子の制御の際に入力されるであろう範囲内で与えた．

- $-1.57 < \theta < 1.57$
- $-3.0 < \dot{\theta} < 3.0$
- $F = -10$ or 10

実験は一回の学習を 1 ステップ，100000 ステップを 1 試行とし，係数の初期値を変えて 10 試行行った．

結果

それぞれの内部モデルでの 10 試行の中央値，最大値，最小値をグラフに表す．なお，100 ステップ間隔で移動平均を取った．

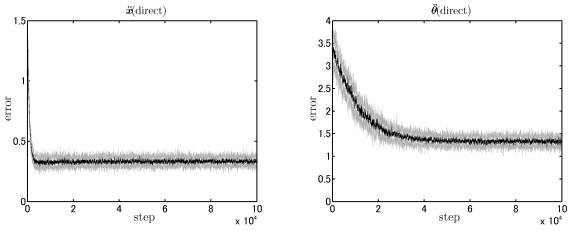
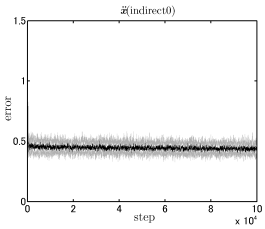


図 5.2　直接報酬のみ



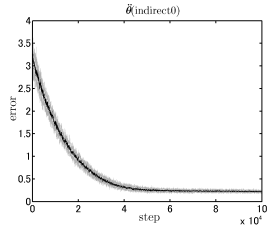
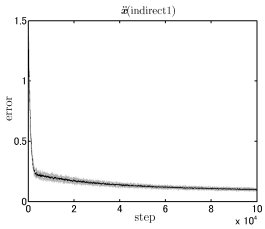


図 5.3　間接報酬パターン 0



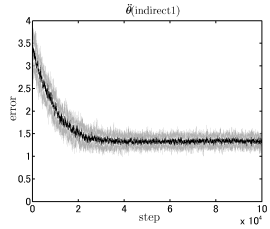
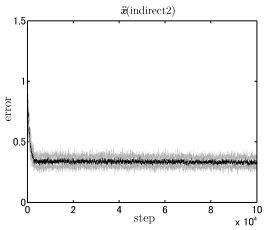


図 5.4　間接報酬パターン 1



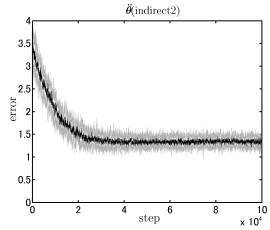
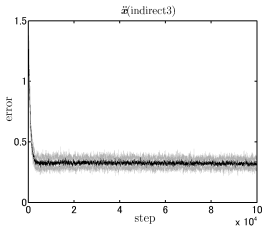


図 5.5　間接報酬パターン 2



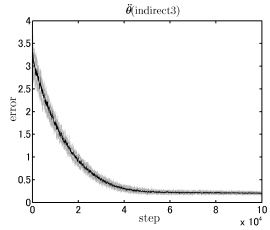
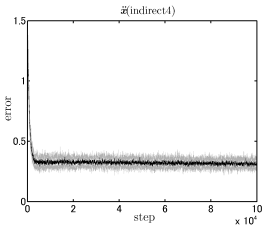


図 5.6　間接報酬パターン 3



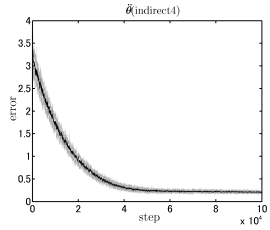
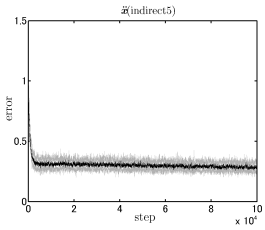


図 5.7　間接報酬パターン 4



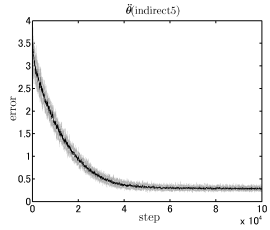
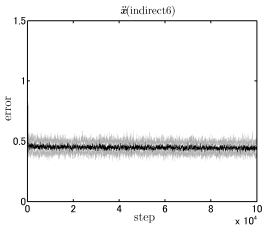


図 5.8　間接報酬パターン 5



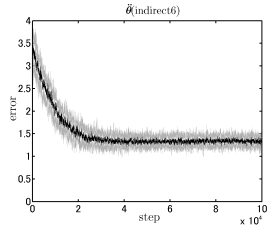
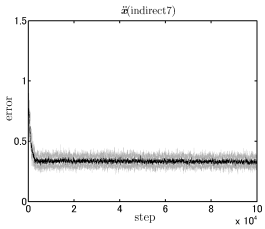


図 5.9　間接報酬パターン 6



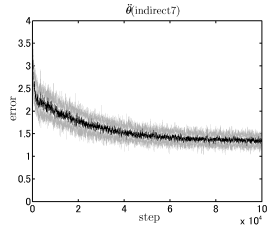
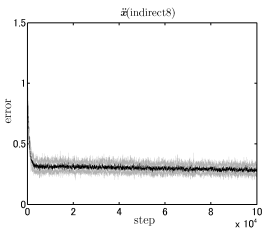


図 5.10　間接報酬パターン 7



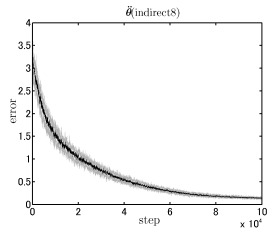
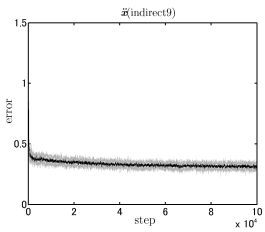


図 5.11　間接報酬パターン 8



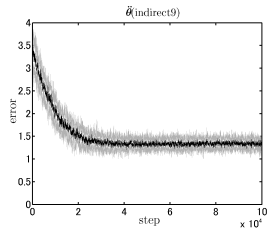
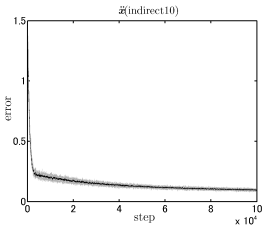


図 5.12　間接報酬パターン 9



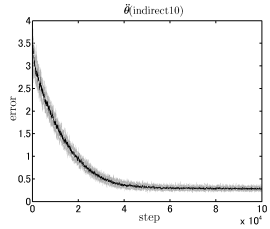
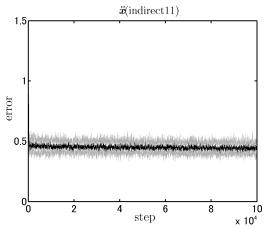


図 5.13　間接報酬パターン 10



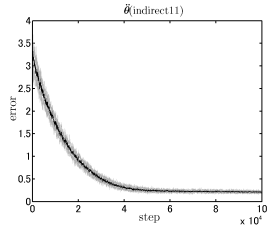
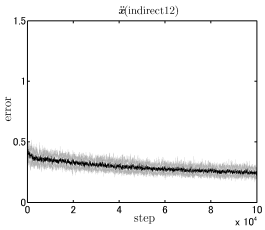


図 5.14　間接報酬パターン 11



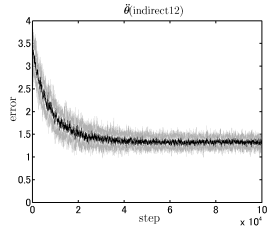
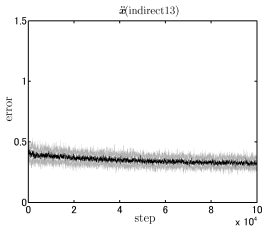


図 5.15　間接報酬パターン 12



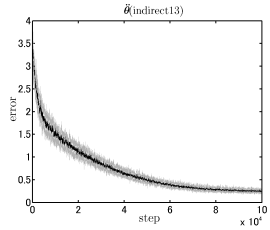
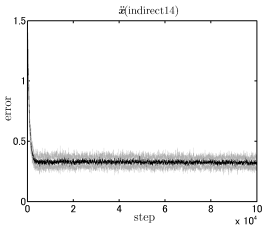


図 5.16　間接報酬パターン 13



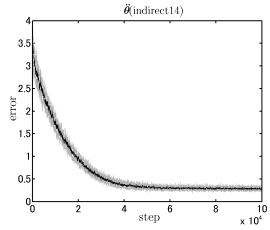
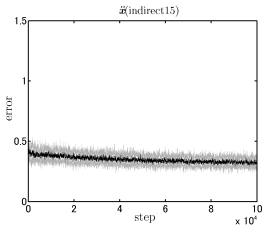


図 5.17　間接報酬パターン 14



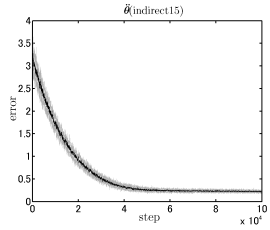
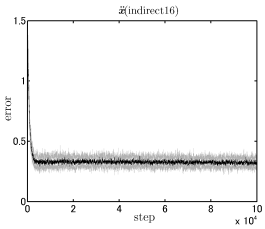


図 5.18　間接報酬パターン 15



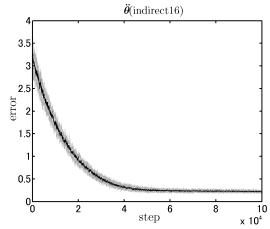
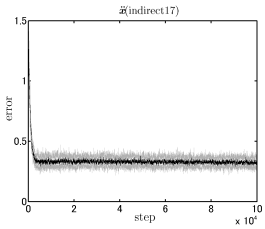


図 5.19　間接報酬パターン 16



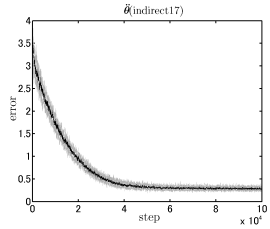
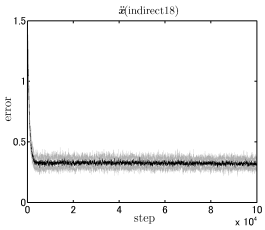


図 5.20　間接報酬パターン 17



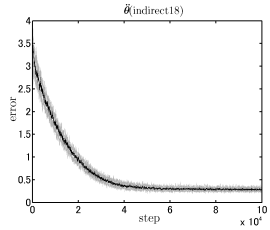
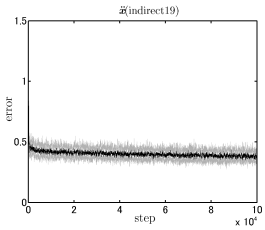


図 5.21　間接報酬パターン 18



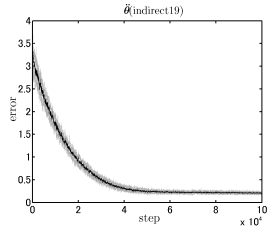
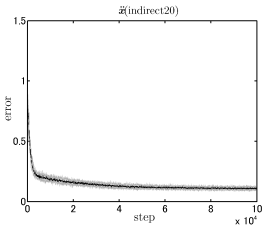


図 5.22　間接報酬パターン 19



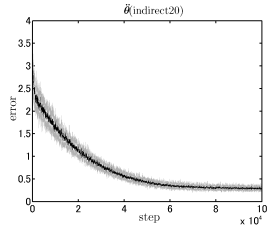
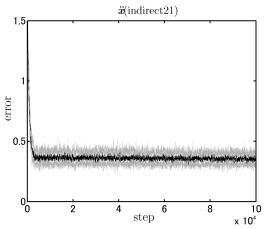


図 5.23　間接報酬パターン 20



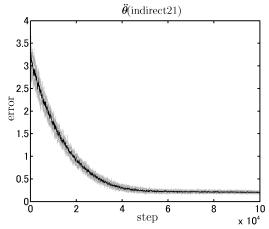
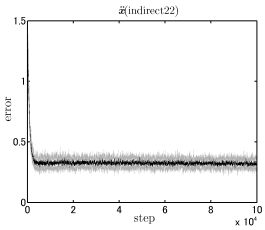


図 5.24　間接報酬パターン 21



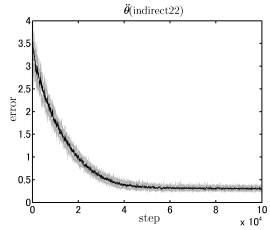
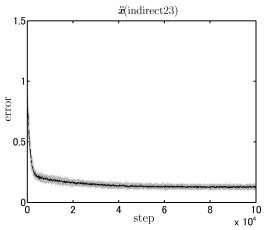


図 5.25　間接報酬パターン 22



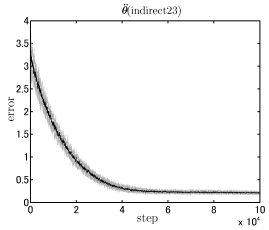
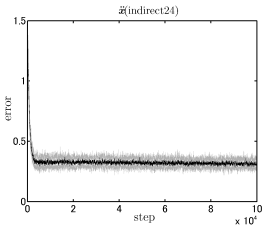


図 5.26　間接報酬パターン 23



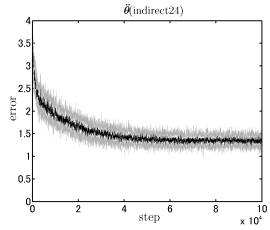
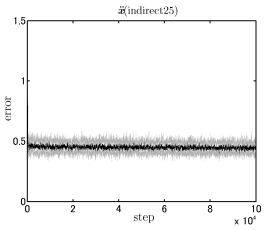


図 5.27　間接報酬パターン 24



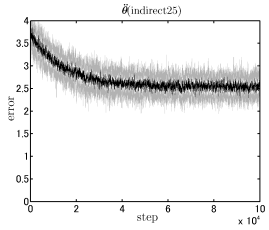
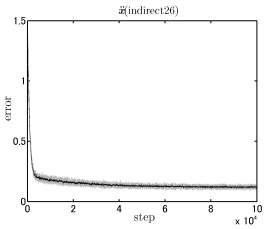


図 5.28　間接報酬パターン 25



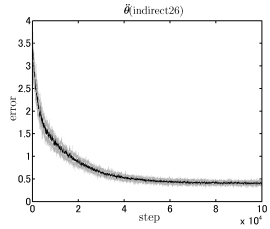
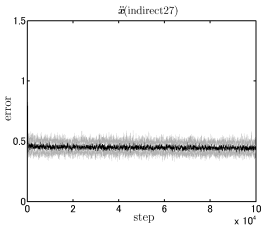


図 5.29　間接報酬パターン 26



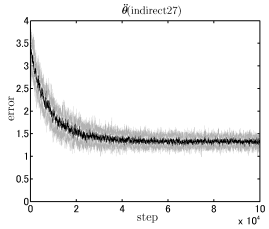
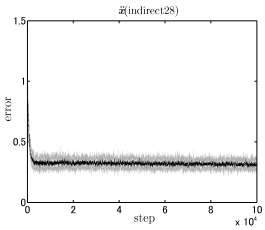


図 5.30　間接報酬パターン 27



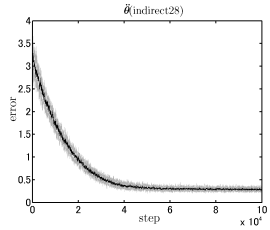
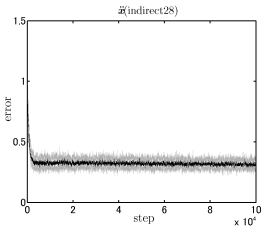


図 5.31　間接報酬パターン 28



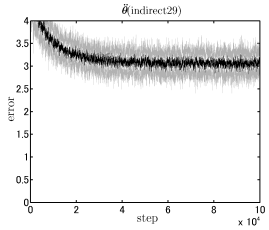


図 5.32　間接報酬パターン 29

グラフを見ると，誤差が収束したときの値から，$\ddot{x}, \ddot{\theta}$ ともに大別して予測精度の高いものと低いものの 2 種類に分けることができる．直接報酬のみ (図 5.2)，パターン 2(図 5.5)，パターン 9(図 5.12)，パターン 24(図 5.27) のように $\ddot{x}, \ddot{\theta}$ 両方の予測精度が低いにも関わらず制御に成功しているエージェントがあることから，これらのパターン程度の予測精度があれば，この環境でのタスクは達成できるようである．しかし，一方ではパターン 23(図 5.26) のように $\ddot{x}, \ddot{\theta}$ ともに精度の高い予測が出来ているにも関わらず，制御が出来ていないエージェントもある．このことから，精度の高い状態予測器を持っていても，行動決定器の性能が良くなければ制御はできないということがわかる．

以下では，さらに細かい分析を行うために，状態予測器の誤差収束性能を評価する値 ϵ と cs を図 5.33 のように定義する．ϵ は最終的に収束した誤差の値，cs は 100000 ステップの学習の結果減少した誤差の 10% を ϵ に加えたものを閾値として，閾値を最初に下回ったときのステップ数である．ϵ，cs ともに小さいエージェントほど性能が良い．

各エージェントの状態予測器の誤差収束性能を表 5.3 に示す．

表 5.3 から $\ddot{x}, \ddot{\theta}$ ともに最も良い内部モデルを持っているのはパターン 10(図 5.13) の状態予測器であることがわかる．

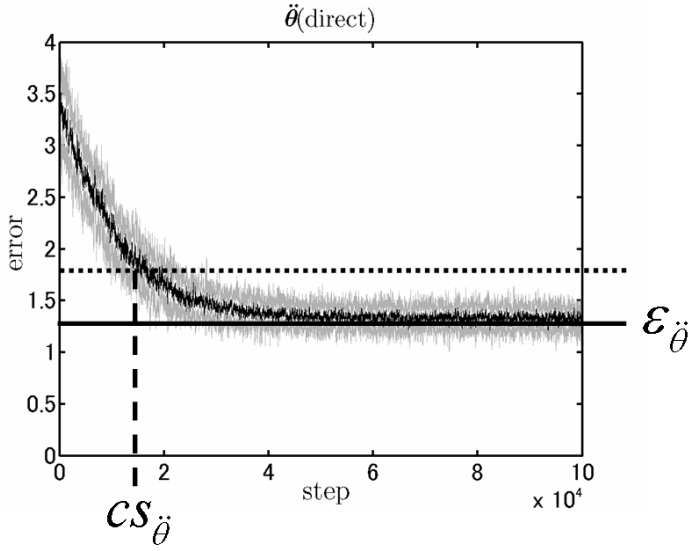


図 5.33　誤差収束性能評価値の例

表 5.3　各状態予測器の誤差収束性能

報酬	$\epsilon_{\ddot{x}}$	$cs_{\ddot{x}}$	$\epsilon_{\ddot{\theta}}$	$cs_{\ddot{\theta}}$
直接報酬のみ	0.3225	678	1.2976	18945
間接報酬 0	0.4349	188	0.2082	26479
間接報酬 1	0.0964	968	1.3200	15707
間接報酬 2	0.3166	204	1.3107	16366
間接報酬 3	0.3062	699	0.1920	27658
間接報酬 4	0.2986	721	0.1926	26481
間接報酬 5	0.2781	242	0.2753	25003
間接報酬 6	0.4407	187	1.2994	16623
間接報酬 7	0.3161	195	1.3411	23191
間接報酬 8	0.2765	244	0.1367	38530
間接報酬 9	0.3195	200	1.2933	16631
間接報酬 10	0.0930	968	0.2759	24997
間接報酬 11	0.4393	189	0.1960	27459
間接報酬 12	0.2411	132	1.3907	13450
間接報酬 13	0.3099	121	0.2483	39260
間接報酬 14	0.3086	700	0.2756	24934
間接報酬 15	0.3099	121	0.2086	26478
間接報酬 16	0.3103	705	0.2078	26480
間接報酬 17	0.3119	698	0.2753	25003
間接報酬 18	0.3073	705	0.2761	25000
間接報酬 19	0.3895	194	0.1912	27069
間接報酬 20	0.1020	595	0.2833	36568
間接報酬 21	0.3400	913	0.1912	27069
間接報酬 22	0.3065	700	0.2958	26502
間接報酬 23	0.1177	588	0.1967	27460
間接報酬 24	0.2980	722	1.3162	18453
間接報酬 25	0.4436	188	2.4764	22798
間接報酬 26	0.1138	942	0.4025	23268
間接報酬 27	0.4403	187	1.3085	14784
間接報酬 28	0.2979	222	0.2788	26080
間接報酬 29	0.2979	222	2.9927	17616

最後に，実際に状態予測器がどのような内部モデルを具体的に持っているのかを確認する．ここでは，直接報酬のみの状態変数の持つ内部モデルと最良の状態予測器としてパターン 10 の持つ内部モデル，そして運動方程式をもとに理想的な素子を選択した準完全内部モデルを示す．なお，a_i，b_i は結合係数を表す．

間接報酬パターン 10

$$\begin{aligned}\ddot{x} &= a_1 \sin\theta + a_2 \sin^2\theta + a_3\dot{\theta}^3 \cos\theta + a_4 F + a_5\dot{\theta}^2 \sin\theta \\ &\qquad + a_6 \sin\theta\cos\theta + a_7 F \sin^2\theta + a_8\dot{\theta}^3 \sin^2\theta \\ \ddot{\theta} &= b_1 \sin\theta + b_2\dot{\theta}^3 \cos\theta + b_3\dot{\theta} + b_4 F + b_5 F \sin^2\theta + b_6 \sin\theta\cos\theta \\ &\qquad + b_7 F \sin\theta\cos\theta + b_8 \cos^3\theta\end{aligned}$$

直接報酬

$$\begin{aligned}\ddot{x} &= a_1 \sin\theta + a_2 \cos\theta + a_3\dot{\theta} + a_4 F + a_5 \sin^2\theta + a_6 F \sin^2\theta \\ &\qquad + a_7 \cos^2\theta + a_8\dot{\theta}^2 \cos\theta \\ \ddot{\theta} &= b_1 \sin\theta + b_2 \cos\theta + b_3\dot{\theta} + b_4 F + b_5\dot{\theta}\cos^2\theta + b_6 \cos^2\theta \\ &\qquad + b_7\dot{\theta}^2 \cos\theta + b_8\dot{\theta}\cos\theta\end{aligned}$$

準完全内部モデル

$$\begin{aligned}\ddot{x} &= a_1 F + a_2\dot{\theta}^2 \sin\theta + a_3 \cos\theta\sin\theta + a_4 F\cos^2\theta + a_5\dot{\theta}^2 \sin\theta\cos^2\theta \\ \ddot{\theta} &= b_1 \sin\theta + b_2 F\cos\theta + b_3\dot{\theta}^2 \sin\theta\cos\theta\end{aligned}$$

倒立振子の運動方程式は準完全内部モデルを見てもわかる通り，$\ddot{x}$ は 5 つの項，$\ddot{\theta}$ は 3 つの項で近似することができる．状態予測器の内部モデルは 8 つの素子でこれらの項を表現できれば良い．間接報酬パターン 10 のエージェントの内部モデルは，$\ddot{x}$ の式では，$F, \dot{\theta}^2\sin\theta, \sin\theta\cos\theta$，$\ddot{\theta}$ の式では $\sin\theta$，の合計 4 つの素子を直接持っている．一方，直接報酬のみのエージェントの内部モデルは $\ddot{x}$ の式では F，$\ddot{\theta}$ の式では $\sin\theta$，の合計 2 つの素子しか持っていない．その他の素子の組み合わせで，それ以外の項を近似していることも考えられるが，この 2 つの内部モデルの誤差収束性能を考えると妥当な結果であると言える．

5.6　実験３：総合的な技能の比較

最後に，状態予測器単体だけではなく，行動決定器の持つ状態価値関数を含めた，エージェントの獲得した総合的な技能についての比較を行う．具体的には環境のパラメータを変えて技能を獲得した際よりも振子の制御が難しいような条件を作り出し，獲得した技能でその条件に対応できるかどうかの検討を行う．

前にも述べたように，簡単な条件下では初級者と上級者のタスクの結果に明示的な差は現れないが，厳しい条件下においては初級者と上級者のタスクの結果は大きく異なる．つまり条件が厳しくなるほどに，その技能の差がタスクの結果として顕著に現れるのである．

今回タスクとして扱った倒立振子の場合，エージェントが台車を押す力に対して台車の重さが小さくなるほど，また振子の長さが大きくなるほど，一度の行動の及ぼす影響が大きくなる．つまり，制御が難しくなると考えられる．そこで，最初に学習を行った実験１の環境よりも厳しい環境を表 5.4 のように設定し，実験１での学習終了直後の状態予測器と行動決定器を使って，計 15 パターンの環境で制御・学習を行わせた．この結果から，エージェントが実験１で獲得した技能で難度の高い環境に対応できるかどうかを観察する．

結果

結果を表 5.5,5.6,5.7 に示す．また，各エージェントの制御成功環境数を図 5.34 に示す．制御成功環境数を見ると 20 個の間接報酬ありのエージェントのうち，11 個のエージェントが直接報酬のみのエージェントよりも多くの環境で制御が可能であった．また，これら 11 個のエージェントのうち，特に制御成功環境数が 10 を超えたパターン 14,16,17,19,21 に関しては，$\ddot{\theta}$ の内部モデルの予測精度が高いという共通点があった．一方で，$\ddot{x}, \ddot{\theta}$ の両方に対して最も精度の高い内部モデルを持っているパターン 10 やそれに次いで，精度の高かったパターン 20 のどちらのエージェントも他の環境への適応性という点では直接報酬のみのエージェントに劣るということがわかった．この結果は，技能は状態予測器だけによるものではなく，行動決定器と状態予測器の両方の性能で決まるものだということを示している．

表 5.4　用意した他環境一覧

l	振子の重心までの距離	1.50 [m] (Short) 1.75 [m] (Half) 2.00 [m] (Long) 2.50 [m] (Extra Long)
M	台車の質量	1.0 [kg] (Heavy) 0.8 [kg] (Middle) 0.6 [kg] (Light) 0.2 [kg] (Extra Light)
m	振子の質量	0.15 [kg] (Const)

	Short	Half	Long	Extra Long
Heavy	学習環境	環境 3	環境 6	環境 10
Middle	環境 1	環境 8	環境 4	環境 12
Light	環境 2	環境 5	環境 7	環境 11
Extra Light	環境 9	環境 13	環境 14	環境 15

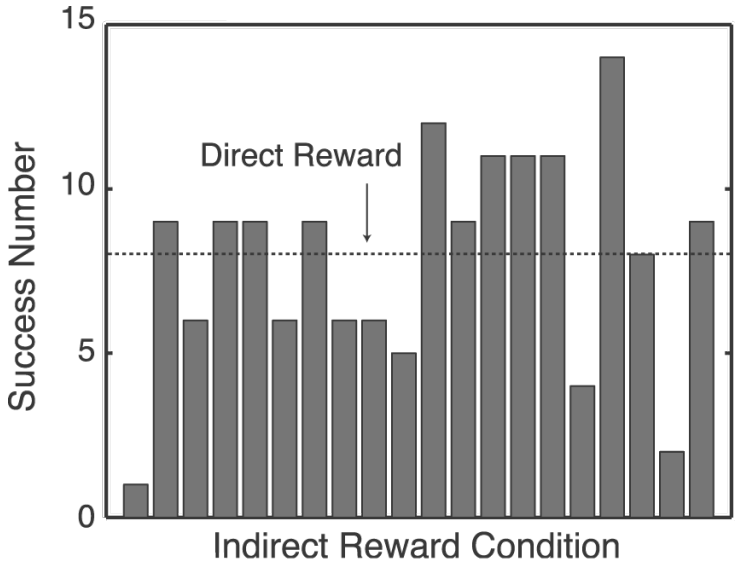


図 5.34　各エージェントの制御成功環境数

表 5.5　他環境でのタスク達成度 1

	環境 1	環境 2	環境 3	環境 4	環境 5
直接報酬のみ	15	制御不可	807	69	20
間接報酬 2	制御不可	3	制御不可	制御不可	制御不可
間接報酬 3	47	47	199	113	59
間接報酬 5	64	15	制御不可	343	12
間接報酬 6	1	1	8	制御不可	643
間接報酬 7	1	10	1	614	13
間接報酬 8	200	28	7	制御不可	制御不可
間接報酬 9	54	12	9	17	43
間接報酬 10	13	13	1364	1644	制御不可
間接報酬 11	20	制御不可	1460	417	制御不可
間接報酬 12	883	133	2207	575	111
間接報酬 14	11	6	122	210	1302
間接報酬 15	69	58	14	336	制御不可
間接報酬 16	31	66	243	590	162
間接報酬 17	1	1	421	279	79
間接報酬 19	4	40	234	2204	80
間接報酬 20	制御不可	制御不可	制御不可	制御不可	制御不可
間接報酬 21	38	9	558	745	55
間接報酬 22	14	18	制御不可	制御不可	1550
間接報酬 24	制御不可	677	制御不可	制御不可	制御不可
間接報酬 27	19	24	37	制御不可	7
制御成功報酬数	18	18	16	14	14

表 5.6　他環境でのタスク達成度 2

	環境 6	環境 7	環境 8	環境 9	環境 10
直接報酬のみ	2672	制御不可	10	制御不可	制御不可
間接報酬 2	制御不可	制御不可	制御不可	制御不可	制御不可
間接報酬 3	329	制御不可	制御不可	75	174
間接報酬 5	制御不可	195	77	制御不可	制御不可
間接報酬 6	2802	制御不可	3	28	制御不可
間接報酬 7	制御不可	37	1	制御不可	3
間接報酬 8	2088	制御不可	制御不可	44	253
間接報酬 9	1050	16	16	制御不可	制御不可
間接報酬 10	711	645	制御不可	制御不可	制御不可
間接報酬 11	2565	制御不可	制御不可	制御不可	1468
間接報酬 12	制御不可	制御不可	制御不可	制御不可	制御不可
間接報酬 14	1007	214	141	制御不可	545
間接報酬 15	1994	926	462	制御不可	287
間接報酬 16	478	504	279	167	1689
間接報酬 17	1226	236	311	87	2040
間接報酬 19	制御不可	168	291	91	488
間接報酬 20	制御不可	制御不可	制御不可	744	制御不可
間接報酬 21	1002	225	372	107	制御不可
間接報酬 22	414	2710	制御不可	92	制御不可
間接報酬 24	制御不可	2236	制御不可	制御不可	制御不可
間接報酬 27	6	349	制御不可	78	制御不可
制御成功報酬数	14	13	11	10	9

表 5.7　他環境でのタスク達成度 3

	環境 11	環境 12	環境 13	環境 14	環境 15	制御成功 環境数
直接報酬のみ	39	950	制御不可	制御不可	制御不可	8
間接報酬 2	制御不可	制御不可	制御不可	制御不可	制御不可	1
間接報酬 3	制御不可	制御不可	113	制御不可	制御不可	9
間接報酬 5	制御不可	制御不可	制御不可	制御不可	制御不可	6
間接報酬 6	426	制御不可	制御不可	705	制御不可	9
間接報酬 7	制御不可	27	制御不可	制御不可	制御不可	9
間接報酬 8	制御不可	制御不可	制御不可	制御不可	制御不可	6
間接報酬 9	23	制御不可	制御不可	制御不可	制御不可	9
間接報酬 10	制御不可	制御不可	制御不可	制御不可	制御不可	6
間接報酬 11	599	制御不可	制御不可	制御不可	制御不可	6
間接報酬 12	制御不可	制御不可	制御不可	制御不可	制御不可	5
間接報酬 14	119	622	制御不可	制御不可	2270	12
間接報酬 15	制御不可	313	制御不可	制御不可	制御不可	9
間接報酬 16	制御不可	740	制御不可	制御不可	制御不可	11
間接報酬 17	制御不可	498	制御不可	制御不可	制御不可	11
間接報酬 19	378	制御不可	44	制御不可	制御不可	11
間接報酬 20	98	制御不可	269	249	制御不可	4
間接報酬 21	1034	1116	151	283	494	14
間接報酬 22	制御不可	制御不可	制御不可	126	327	8
間接報酬 24	制御不可	制御不可	制御不可	制御不可	制御不可	2
間接報酬 27	制御不可	制御不可	69	163	制御不可	9
制御成功 報酬数	8	7	5	5	3	165

5.7　まとめと考察

実験１の結果，間接報酬はただ付加すればいいわけではなく適切な値を付加しなければ逆効果であることがわかった．このことは，実際の運動にも言えることで，間違ったコツや癖を身に付けてしまうと逆に学習の妨げになってしまう場合と同様であると考えられる．

実験２では，実験１で獲得した技能のうち，状態予測器の内部モデルの性能を見た．20個の間接報酬ありのエージェントの中から最も性能の良いものと直接報酬ありのエージェントの内部モデルを比較し，間接報酬ありのエージェントの持つ内部モデルが倒立振子の運動方程式に近いことを確認した．

実験３では，実験１で獲得した総合的な技能を評価するために，学習で用いた環境とは別の未経験の環境での倒立振子の制御を行った．「学習した知識を活かして，異なった環境へ適応する」というのは第２章で述べた上級者の定義である．学習環境以外で制御を行うためには，状態予測器と行動決定器の両方が環境に依存しない普遍的な知識としてエージェントに身に付いていなければならない．

実験の結果，間接報酬ありのエージェントのうち，11 個のエージェントが直接報酬のエージェントよりも多くの環境で制御が可能であった．このことから，直接報酬のみのエージェントよりもこれらの間接報酬ありのエージェントの方が高い技能を持っているということがいえる．

以上の結果から，間接報酬を用いることでこのモデル上で初級者と上級者の技能の差を表現することができたといえる．

第 6 章

議論

ここでは，ここまでのモデル上では議論できなかった問題について考察する．

6.1 “適切な” 間接報酬とは

第５章での実験の結果，“適切な” な間接報酬を与えることで学習者の技能を向上させられることがわかった．ここではこの適切な間接報酬について考察する．

前にも述べたように，間接報酬とは，熟練経験者による教示や他者の運動観察にもとづく評価などから得られる「コツ」に相当する．例えば，スキー初心者が転んでしまったとき，板の先を掴むようにして起き上がると楽に起き上がることができる.「コツ」とはこのような，その状態や動作を経由することで目標の運動の生成を容易にするものである．

一般に，ある一つの運動に関する「コツ」には, 多種多様なものが存在するが，その全てが全ての学習者に対して一様に効果を発揮するわけではない．これは学習者の持つ内部モデルに個人差があり，それを修正するための「コツ」も人それぞれだからである．例えば，悪いクセをつけてしまった学習者に，そのクセを修正させるためには，特別な（パフォーマンスが一時的に下がってしまうような）教示が必要になることもある．これは，すでに獲得した内部表現構造や行動選択ルールを壊して，適切な技能の再構成を促すための手段であると考えられる．

また, 教示や他者の運動の観察からどのような「コツ」を得るかも学習者によって異なる．例えば，学習者がどこに注意を向けるか，受けた教示をどのように解釈するかによって，同じ運動を観察しても，また同じ内容の教示を受けても，学習者が得る「コツ」は異なり，場合によっては，教示者の意図とは違う「コツ」を得ていることも考えられる．

このように考えると，適切な間接報酬は学習者の持つ内部モデルや行動選択ルールによって異なると言える．つまり，ある学習者に有効な間接報酬が別の学習者にも有効だとは限らない．だからこそ，効率的な運動学習のためには，学習者に合わせた間接報酬を与

えることのできる優秀な教示者が必要なのである．

学習者がどのような内部モデルや行動選択ルールを持っているかは他者はもちろん学習者自身も正確に把握することは困難である．また，その運動の結果が何に起因しているのか，内部モデルに問題があるのか，そもそも行動選択ルールの基となる観察イメージが間違っているのか，を把握することも同様に困難である．教示者に与えられた「コツ」を学習者が実際にどのように解釈しているかもわからない場合もある．

このように考えると，優秀な教示者とは，表にあらわれない学習者の間接報酬や内部モデルを，運動の様子から類推し，その修正のために適切な間接報酬を，意図した通りに適時与えることのできる者だと考えられる．

6.2 素子の作成・追加

第４章のシミュレーションモデル上では，状態予測器の内部モデルの特徴素子は，ある一定条件を満たして学習が停止すると機械的に追加されていた．そのため，直接報酬しかもたないエージェントでも，タスク達成後も引き続き学習を行うことで，より良い素子を獲得できるようになる場合もあり得る．しかし，実際の運動学習では直接報酬しかもたないエージェント，つまりタスク達成しか考えていない学習者が，新たな素子を作成する，生成できる運動の次元を増加させることはできないと考えられる．なぜなら，実際の運動の次元は機械的に増加するようなものではなく，試行錯誤による “産みの苦しみ” を要するからである．こうした試行錯誤では今までには行わなかった運動を意識して行わなければならないため，新たな運動の次元を探索する最中にはパフォーマンスが落ちる．そのため，タスク達成だけを目的に運動を行う学習者はそのような運動を避ける．よって，どれだけ学習時間を長くとっても，こういった学習者が新たな運動の次元に気付くことはない．

実際の運動学習ではこうした試行錯誤を効率良く行うトレーニングとして，バリエーショントレーニングというものが広く知られている．バリエーショントレーニングとは，意図的に学習者の運動を制限することで，学習者の探索空間を狭め，新たな運動の次元に気付きやすくするトレーニングのことである．

最終的に目標とする運動とは異なる運動を学習させることにより，通常の学習過程では経験しづらい状態と，今までに形成してきた内部モデルでは対応できない環境をあえて経験させ，今までには無かった素子の作成，つまり内部モデルの学習，を促進させることができると考えられる．

第 7 章

結論

本研究では，運動技能のうちの内的過程に焦点をあて，人が学習を行うことによって何が変化するのか，どのように初級者から上級者へと変化していくのかについて論じた．

我々人間は「行おうとしている運動に関する情報」をイメージという形で持っており，試行錯誤や見まねや上級者による教示を通じて，イメージを変化させていく．内的過程の役割とはこのイメージを具体的な運動指令へと変換することである．

理論モデルでは，体性感覚や視覚情報など自らの中だけの情報を内的フィードバック，他者からの教示や手本の観察，運動の結果に関する外界からの評価などを外的フィードバックと学習に必要なフィードバック情報を２種類に分けて，上述の内的過程の役割を簡略化したモデルを提案した．

シミュレーションモデルでは,「行おうとしている運動に関する情報」を状態価値関数として持った行動決定器と，試行錯誤を通じてイメージ通りの動きを生成するための機構として状態予測器を持った強化学習ベースのモデルを提案した．理論モデルの内的フィードバックは状態予測器に返される予測誤差，外的フィードバックは報酬という形で実装した．同じ運動を学習する際でも，外的フィードバックはその学習環境によって与えられるものが異なる．例えば，同じ運動を学習する際でも，コーチが付いているかいないかによって受けるフィードバックは異なる．そこで，外的フィードバックに対応する報酬を直接報酬と間接報酬という２種類に分けることでこの異なるフィードバックを表現した．

最後に実験では，計算モデル上で，直接報酬のみで学習するエージェントと間接報酬ありで学習するエージェントに同じタスクを学習させ，その状態予測器と行動決定器の差，つまり技能の差を検討した．最終的には適切な間接報酬を用いることで初級者と上級者の技能の差，また初心者が初級者もしくは上級者へと技能を変化させていく過程を表現することができた．

謝辞

本研究を進めるにあたり，御多忙の中，最後まで熱心な御指導，御支援を賜りました阪口 豊 助教授と出澤 正徳 教授に深く感謝するとともに，ここに厚く御礼申し上げます。同様に，石田 文彦 助手，島井 博行 助手には，本研究が緒についたおりから，終始有益な御討論および適切な御教示を賜りました．ここに，謹んで感謝の意を表します．

最後に，ゼミ発表などでご意見や激励を下さったヒューマンインターフェース学講座の皆様に，いつも暖かく見守ってくれた家族に，心より感謝申し上げます．

参考文献

[1] 神宮: “スキルの認知心理学”, 川島書店, 1993.

[2] 辻本, 竹内:“技能修得過程で用いられる言語的教示と身体動作の認知的連関”, ヒューマンインターフェース学会研究報告集, Vol.4, No.1,pp. 1-6, 2002

[3] 浅岡:“動きの模倣とイメージトレーニング”, バイオメカニズム学会誌, Vol.29, No.1, pp31-35, 2005.

[4] 長谷川, 星野:“スポーツ選手のスキルと身体運動イメージの関係”, 順天堂大学スポーツ健康科学研究 Vol.6, pp. 166-173, 2002.

[5] 佐々木:“からだ：認識の原点”, 東京大学出版会, 1987.

[6] 田口, 阪口, 島井, 石田:“運動観察時の注視点の動き”, 電気通信大学大学院 IS シンポジウム “Sensing and Perception” 第 13 回予稿集,pp.5-8, 2006.

[7] 阪口, 中野:“ボトムアップ型学習による階層ネットワークの自己組織化”, 計測自動制御学会, July, pp 22-24, 1992.

[8] “Adaptive critics and the basal ganglia”, Models of Information Processing in the Basal Ganglia, pp. 215-232, MIT Press, 1994.

[9] 鮫島, 銅谷: “強化学習と大脳基底核”, バイオメカニズム学会誌, Vol.25, No.4, pp. 167-171, 2001.