C-07

モジュール組換え型アーキテクチャを用いた行動学習法の検討

Action Learning by an Agent with Module Recombination Architecture

坂本 裕太†　本多 透†　尾関 基行†　岡 夏樹†
Yuta Sakamoto　Toru Honda　Motoyuki Ozeki　Natsuki Oka

1．はじめに

問題解決を行う知的システムにおいて，行動がすべて作り込まれたロボットでは，設計者が想定した問題にしか対応できず柔軟性に欠ける．そのため，変動する環境に対応させようと，ロボットに環境と行動の関係を学習させるさまざまな手法が提案されてきた．その中でも人間の脳のモジュール性に注目した研究がある[1]．人間の脳は各領域が異なる機能を持ち，それらを組み合わせることで複雑な問題に素早く適応している．ロボットの各機能も同様にモジュール化し，それらを組み換えることで柔軟に問題を解決するシステムを構築できる可能性がある．

このようなモジュール型学習の利点は複数ある．まず，複雑な学習問題でも，より簡単な下位の問題に切り分けてモジュールごとに学習させることで対応できる．また，モジュール同士のインターフェースを固定しておけば，モジュールを取り換えることで学習をし直さずに異なる動きをさせることができる．さらに，異なる環境でも，モジュールの組み換え方を学習させることで，ロボットが自律的に行動を獲得することができる．

しかし，モジュール型学習には，モジュール数の増加に伴い，組み合わせの数が爆発するという問題がある．また，学習の際の計算量はさらに環境の状態数に比例して増加する．そのため，実用化するためには組み合わせを限定することなどが必要である．

この問題に対し，本多らは，モジュールの組み換えの学習に言語教示を援用する手法を提案している[2]（モジュール組み換え型アーキテクチャ[3]）．本多らの研究では，モジュールの組み換えの系列を強化学習により学習する．教示を用いることで，組み合わせの数が爆発するという問題を回避している．教示を用いる利点は，問題をうまく分割するような教示（たとえば音に注意と光に注目）を用いることで学習空間を小さくできることである．ただし，本多らの研究ではモジュールの機能についてはあらかじめ適切なものが与えられていることを前提としており，組み換えに関する学習のみを行っていた．

そこで本研究では，本多らの研究を発展させ，あらかじめ作り込まれた固定モジュールと学習を行うモジュールから構成されるモジュール型システムにおいて，モジュールの組み合わせ方の学習を，モジュールの学習と同時並行で行う同時学習に取り組む．モジュールの機能の学習と組み換えの学習は，両方とも強化学習で行う．本研究では，機能が未学習であるモジュールの組み換えが，うまく学習できるかを実験的に確認する．

本論文では，まず 2 章で，モジュール組み換え型アーキテクチャのモデルと，モジュールの学習とモジュール組み換えの学習について説明し，3 章で実験方法について説明する．4 章では，実験の結果とその考察を行い，5 章で今後の展望について述べる．

†京都工芸繊維大学, Kyoto Institute of Technology

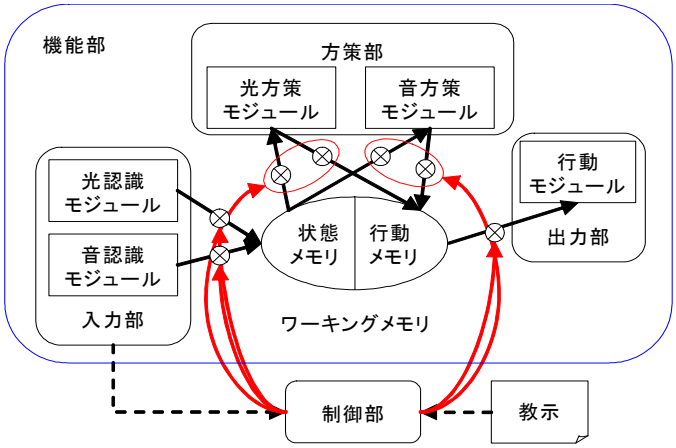


図 1. モジュール組み換え型アーキテクチャ

2．モジュール組み換え型アーキテクチャ

図 1にモジュール組み換え型アーキテクチャの構成を示す．

本アーキテクチャは，大きく制御部と機能部に分かれる．機能部は任意の個数のモジュールから構成されるが，本論文では機能部は，入力部と方策部にそれぞれ 2 つ，出力部に 1 つの合計 5 つのモジュールからなるとする．それぞれのモジュールは，ワーキングメモリを介してつながっている．

制御部は，環境情報と現在のモジュール間結合の on-off の状態，および，与えられる教示を入力として，各モジュール結合を制御する信号を機能部に出力する．機能部は，各モジュール結合が on にされることで，ワーキングメモリを介してモジュールが機能する．具体的には，光方策モジュールを起動する場合，状態メモリと光方策モジュールの結合と光方策モジュールと行動メモリの結合を on にすることで，”状態メモリ→光方策モジュール→行動メモリ”という情報の流れを実現し，光方策モジュールから取るべき行動を受け取ることができる．以降，場面ごとに on にするモジュールを制御信号によって変えていくことをモジュールの組み換えと呼ぶ．ワーキングメモリは，状態メモリと行動メモリからなる．入力部は，モジュールごとに対応したセンサーの値を獲得する．センサー値は状態メモリに渡す．方策部は，状態メモリの値を読み取って，そこからある方策に従って取るべき行動を決定し，行動メモリに渡す．出力部は，行動メモリの値を読み取ってエージェントの行動を出力する．

制御部により，各モジュール間結合の on-off（ワーキングメモリに情報を書き込む／読み込む）を切り替えていくことによって，このアーキテクチャを実装したシステムは，環境情報を得て，方策に従い行動することができる．

2.1 各モジュールの機能

本研究で行う実験では，計算機上で簡単な迷路のシミュレータを作成し，ロボットが障害物を避けながらゴールへ向かう探索タスクを使用した．以下，ゴールから光が発せられ，障害物から音が発せられるとして，各モジュールの機能を示す．

(1) 入力部

エージェントは，エージェント本体の前後左右に備え付けられた2種類のセンサーからのセンサー値を認識する．

- 光認識モジュール：光源の方向や距離を反映した光センサー値を受けとるモジュール．
- 音認識モジュール：音源の方向や距離を反映した音センサー値を受けとるモジュール．

(2) 方策部

強化学習により学習を行い，エージェントが取るべき行動を選択する．

- 光方策モジュール：状態メモリにある情報に基づいて光源に向かう行動を出力する．
- 音方策モジュール：状態メモリにある情報に基づいて音源を避ける行動を出力する．

(3) 出力部

行動メモリにある情報に基づいてエージェントを動かす．

(4) ワーキングメモリ

次のような制約をもたせた．行動メモリに情報がない状態で出力部が起動されると，出力部が起動されてもエージェントは行動しない．状態メモリに情報がない状態で方策部が起動されると，方策部はとるべき行動を返さない．エージェントが行動した場合，ワーキングメモリに記憶されている情報はクリアされる．行動メモリに情報がある状態で方策部が起動されると，行動メモリの情報は上書きされる．

2.2 学習方法

本研究では，強化学習にSarsa(λ)[4]を，行動選択にε-greedy 法を用いる．

(1) モジュール組み換えの学習

モジュールの組み換えの学習は，制御部において行う．モジュールの組み換えの学習では，光認識モジュール，音認識モジュール，光方策モジュール，音方策モジュール，出力部の5つのモジュールの組み換えを行う．また，今回の実験では，モジュールの起動はすべて単体のみで，モジュールの同時複数起動は行わないものとする．“光認識モジュール→状態メモリ→光方策モジュール→行動メモリ→行動モジュール”や“音認識モジュール→状態メモリ→音方策モジュール→行動メモリ→行動モジュール”というように，与えられたタスクに応じて適切なモジュールの組み換えを学習する．制御部は，現状態（与えられた教示と現在のモジュール間結合の状態）において，各行動（次の時点のモジュール間結合の切り換え）をとることの価値を，強化学習により獲得する．

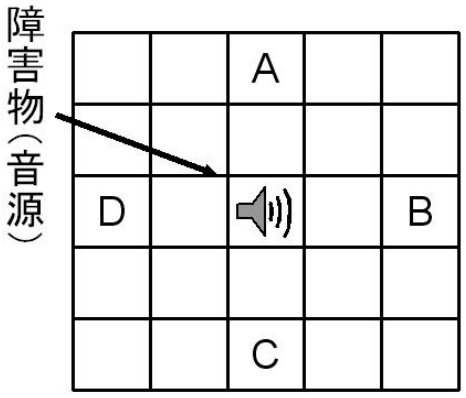


図 2. シミュレーション空間

(2) モジュールの機能の学習

モジュールの機能の学習は，方策部において行う．光方策部では光の情報を，音方策部では音の情報を学習に使うことができる．タイルコーディング[4]を用いて，光または音の認識モジュールから得たセンサー値において，行動（前，後，左，右への移動）を取ったときに次に予測されるセンサー値の状態価値について学習を行う．これによって，方策部は光源に向かう行動または音源を避ける行動を獲得できると考えられる．

3. 実験

3.1 実験タスク

今回のシミュレーションでは，図 2で示すような 5×5 マスの平面上で実験を行った．エージェントは，障害物である音源をよけて，ゴールである光源にたどり着くことが目的である．図 2の A～D の位置に，ゴールである光源とエージェントの初期位置をランダムに設定する．これは，ゴールとエージェントの初期位置を固定することで，エージェントのとるべき行動が固定化され，音を避けるなどの行動を取らなくなることを避けるためである．

エージェントは上下左右に 1 マスずつ進むことができる．エージェントは，位置に応じて光を見て動いて欲しい場面と音を聞いて動いて欲しい場面で異なる教示を与えられながら移動し，障害物のあるマスか，ゴールのあるマスに侵入するまでを 1 試行とする．

3.2 制御部に与える教示について

今回の実験では，光源の上下左右 1 マス隣にエージェントがいる場合，ゴールする可能性が高く，光に注目して欲しいので「光に注目して動いて」という教示を与える．また，音源の上下左右 1 マス隣にエージェントがいる場合，障害物にぶつかる可能性が高く，音に注目して欲しいので「音に注目して動いて」という教示を与える．以降，光に注目して欲しい場合の教示を光教示，音に注目して欲しい教示を音教示と呼ぶ．光教示と音教示のどちらも与えられる場面では，光教示のみ与えることとし，光教示と音教示のどちらも与えられない場面では，教示なしとする．

教示を用いることで，各状況で使用する方策部の使い分けが学習されやすくなることを期待する．

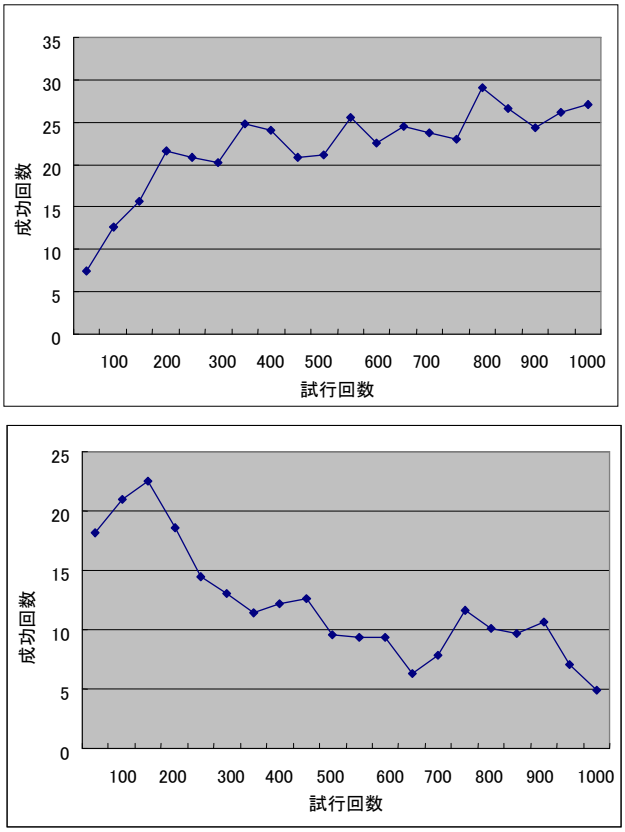


図 3. 成功回数（上：音方策部学習、下：光方策部学習）

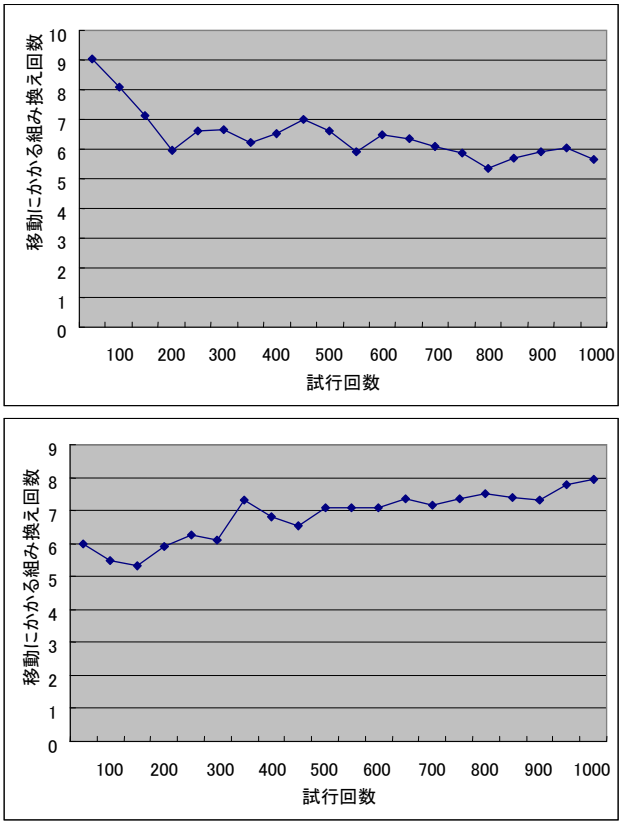


図 4. 1 回の移動にかかる組み換え回数（上：音方策部学習、下：光方策部学習）

3.3 固定する方策部について

本研究では，学習を行うモジュールとあらかじめ作り込まれた固定モジュールを併用する．音方策部で学習を行う場合には光方策部を固定し，光方策部で学習を行う場合には音方策部を固定する．固定された方策部の振る舞いについて，光方策部の場合は，光認識モジュール 4 つのセンサー値で一番大きい値を観測した方向に向かう行動をする．音方策部の場合は，音認識モジュール 4 つのセンサー値で一番大きい値を観測した方向以外に向かう行動をする．

3.4 与える報酬について

方策部には，ゴールした時に大きい正の報酬，障害物にぶつかったときに大きい負の報酬，1 回移動するごとに小さい負の報酬が与えられる．また，エージェントがマスの外に出たときは，小さい負の報酬を与える．制御部には方策部と同様に，ゴールした時に大きい報酬，障害物にぶつかったときに大きい負の報酬が与えられる．また一度のモジュールの切り替えごとに小さい負の報酬が与えられる．

報酬の値については，制御部には，2 つの方策部を固定した状態で，組み換え学習を行ったときによい結果が得られた値を用い，方策部には，一方の方策部を固定し制御部を，”光認識モジュール→状態メモリ→光方策モジュール→行動メモリ→行動モジュール”のように各方策部にそれぞれの状態値を正しい流れで入力するように固定したときによい結果が得られた値を用いた．

4. 実験結果と考察

4.1 実験結果

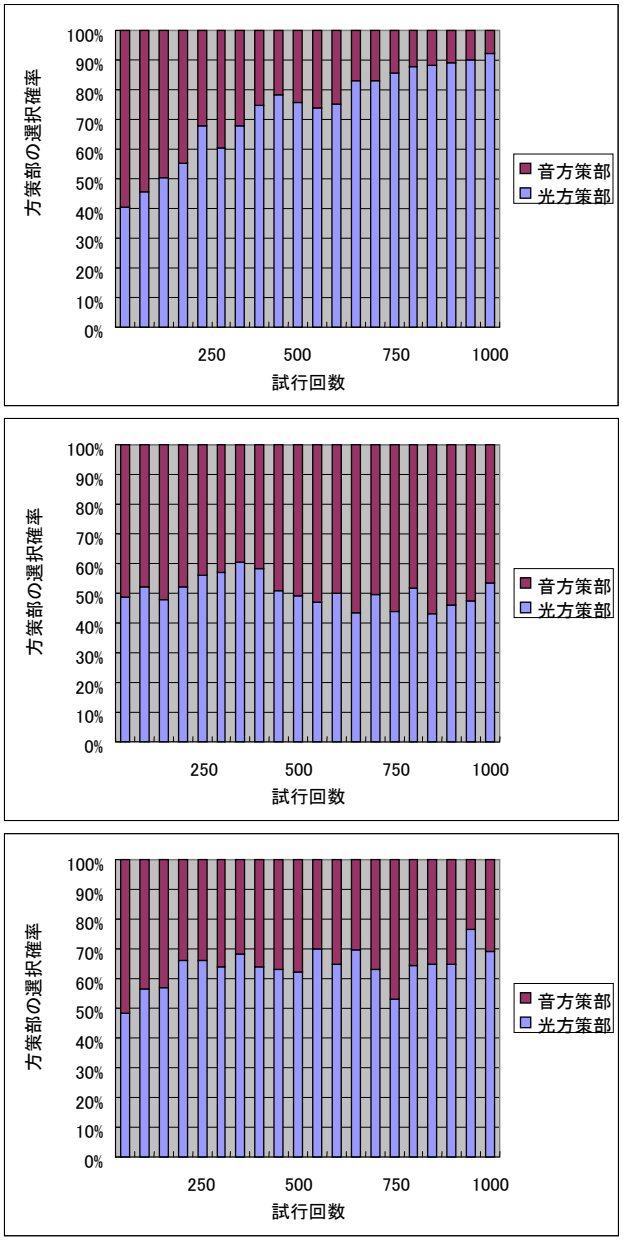
本研究では，モジュールの学習とモジュールの組み合わせの同時学習が可能かを実験的に確認するために，3 章で説明したタスクを 1000 試行行いエージェントに学習させた．そのときの 50 エピソードごとのタスク成功回数，1 回の移動にかかる組み換え回数，各教示でどちらの方策部を選択しているのかを図 3，図 4，図 5に示す．

4.2 考察

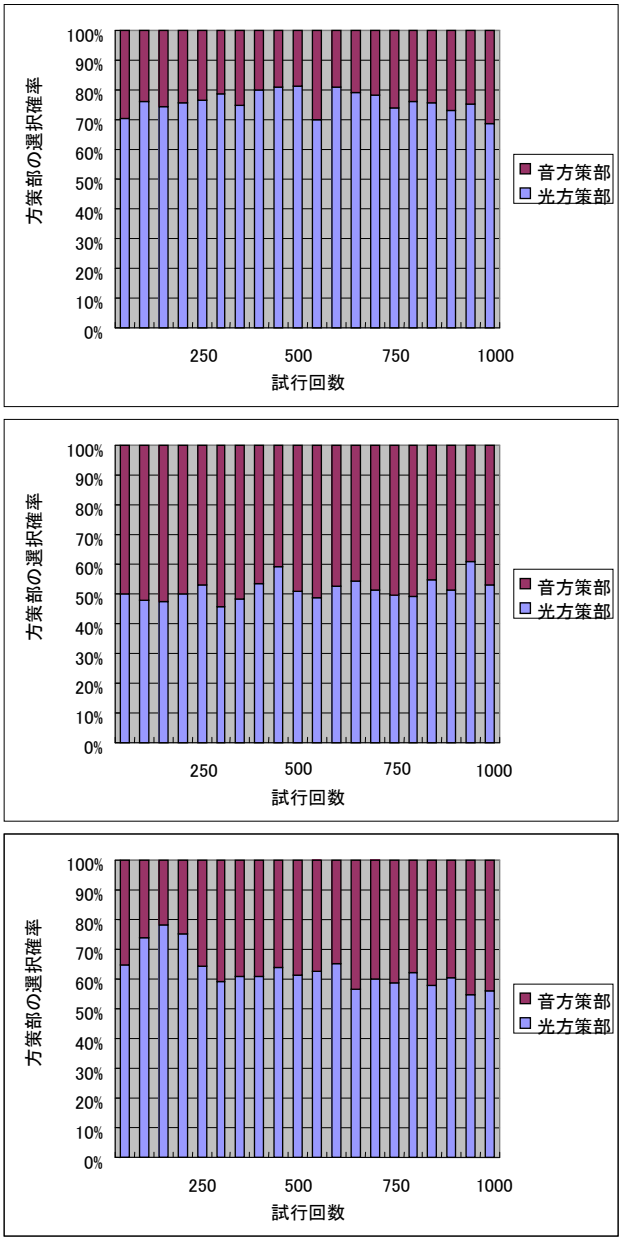
(1) 音方策部学習，光方策部固定の場合

図 3のタスク成功回数と，図 4の 1 回の移動にかかる組み換え回数より，組み換え学習について学習が進んでいることがわかる．また，図 5 (a)の教示に対する方策部の選択について見てみると，ゴールに向かう行動をして欲しい光教示や教示なしのときに光方策部を選択していることが分かる．一方，避けるような行動をして欲しい音教示の場面では，選択にあまり大きな差は出なかった．

また，音方策部の出力する行動について学習できているか調べたところ音源を避けるような行動を行っていた．これより，方策部の学習について問題はなかったと思われる．



(a)．音方策部学習のとき（上:光教示のとき，中:音教示のとき，下:教示なしのとき）



(b)．光方策部学習のとき（上:光教示のとき，中:音教示のとき，下:教示なしのとき）

図 5．音方策部学習時の各教示での方策部の選択確率

(2) 音方策部固定，光方策部学習の場合

図 3，図 4より，組み換え学習とモジュールの学習が進んでいないことがわかる．また，図 5 (b)の教示に対する方策部の選択について見てみると，ゴールに向かう行動をして欲しい光教示のときに光方策部を選択していることがわかるが，教示なしのときには選択にあまり大きな差は出なかった．また，避けるような行動をして欲しい音教示の場面でも，選択にあまり大きな差は出なかった．

そこで，光方策部の出力について学習できているか調べたところ，光源に向かう行動を取るときもあるが，まったく別な，光源に向かわない行動を取ることもあった．これより，方策部の学習について何かしらの問題があったと考えられる．

(3) (1)，(2)の考察

(1)，(2)より，今回の実験ではモジュールとその組み合わせの同時学習について，音教示では期待した教示に関する方策部の使い分けが得られなかった理由として，

- 障害物（音源）が 1 つしかない単純なタスクであったために，エージェントは音教示を与えられた場合でも，光方策部を使って障害物を避けることができ，さらにゴールへ向かう行動ができた

と考えられる．

また，(2)について，タスクの成功回数が低かった理由として，

・　光方策部に与える報酬が適切ではなかった
・　音教示のときに光方策部を用いることで，光方策部の学習がうまく進まなかった

と考えられる．

5. おわりに

本論文では，1)モジュールの機能の学習と，2)その組み合わせ方についての学習- 学習について実験を行った．教示を用いることでモジュールの使い分けを期待していたが，光教示のときには光方策部を用いることが概ね学習されていた．一方，音教示のときには音方策部が用いられるとは限らなかった．これは，タスク設定や報酬の設定に依存して，光方策部の光に向かう行動が音教示の際にも有効であったものと推測される．今後の課題として，適切なタスクや報酬の設定があげられる．また，現在は，光（音）の状態値を扱うものを光方策部（音方策部）としているが，光と音両方の状態値を扱えるような方策部を用いて学習を行うことを計画している．

謝辞

本研究は科研費(21500137) の助成を受けたものである．

参考文献

[1] 小川昭利，大森隆司，“機能部品組み合わせモデルによるナビゲーション行動学習処理の獲得方式の提案，” 電子情報通信学会論文誌，vol.J87-D-II，no.4，pp.987- 998，April 2004.
[2] 本多透，板舛尚樹，岡夏樹，“ロボットの内部情報処理に対する言語教示可能性，”第23 回人工知能学会全国大会論文集，2009.
[3] N. Oka，“‘‘Apparent free will’’ caused by representation of module control,’’ No matter，Never mind: Proceedings of Toward a Science of Consciousness: Fundamental Approaches，pp.243-249，2002.
[4] R.S. Sutton and A.G. Barto，強化学習，三上貞芳，皆川雅章（訳），森北出版株式会社，東京，2000.