3M-08

複合的知識獲得モデルへの並列学習の導入

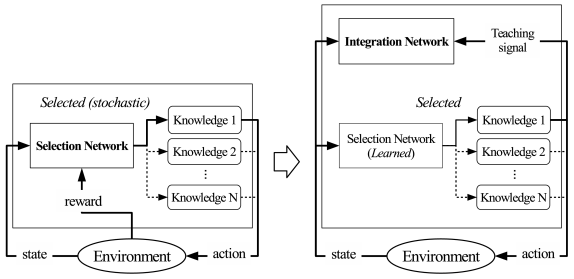
辺見 航平† 服部 元信‡
山梨大学 大学院医工農学総合教育部† 山梨大学 大学院総合研究部‡

1. はじめに

近年,計算機モデルによって知識を獲得する研究が盛んに行われている.中でも矢部[1],識名[2]らが提案した複合的知識獲得モデルは,人間の行動獲得の過程を模した学習モデルである.人間の複合的な知識の獲得例として「ドリブル」という動作の獲得について考える.初めはボールが足元にあるとき は「ボールを蹴る」動作を選択し,ボールが遠くにある場合は「走る」動作を選択する.また,その行動を繰り返し行動することによってぎこちない動作から「ドリブル」という滑らかな動作を徐々に獲得することができる.これは,「ボールを蹴る」,「走る」といった知識を統合し,「ドリブル」という新たな知識を獲得したとみなすことができる.本研究では複合的知識獲得モデルの複合的な知識の獲得過程を改良することによるモデルの効率化,より自然な統合知識の獲得を目指すことを目的とする.以下,複合的知識を統合知識と表現する.

2. 複合的知識獲得モデル

複合的知識獲得モデル[2]とは,ニューラルネットワークを用いた強化学習を行うことにより統合知識を獲得する学習モデルである.複合的知識獲得モデルの構造を図1に示す.



(a) 選択規則学習 (b) 統合知識学習

図 1: 複合的知識獲得モデル

既存知識をあらかじめいくつか与え,タスク環境内で試行錯誤しながら繰り返し行動することにより,それらの既存知識を選択する規則を学習する(a).選択規則を学習する選択ネットワークは,知識を選択する Actor ネットワークとその環境状態の評価値を出力する Critic ネットワークの 2 つのネットワークからなる.どちらのネットワークも環境状態を入力とし,誤差逆伝播法によって学習する.状態遷移の良し悪しは Temporal Difference(TD)誤差を算出することによって判断される.TD 誤差を式(1)に, Critic ネットワークの教師信号を式(2)に示す.

$$td_t = r_t + \gamma c(\mathbf{s}_t) - c(\mathbf{s}_{t-1}) \quad (1)$$

$$c^{teacher}(\mathbf{s}_{t-1}) = c(\mathbf{s}_{t-1}) + td_t \quad (2)$$

式(1)は時刻 $t-1$から時刻 tへの遷移の評価を行う値, r_tは時刻 tでの報酬, γは割引率, $c(\mathbf{s}_t)$は時刻 tの環境状態での Critic ネットワーク出力である.

選択ネットワークの学習完了後,同じ環境下で統合ネットワークの学習を行う(b).統合ネットワークは環境状態を入力し,選択された知識による行動を教師信号として誤差逆伝播法によって学習する.これによって既存知識を統合した知識を獲得することができるモデルとなっている.しかし,このモデルは選択規則の学習終了後,統合ネットワークの学習を行うという逐次的な学習を行う必要があり,非効率であるという問題点が挙げられる.また,人間の知識獲得を模倣したモデルという観点からは,逐次的な学習は不自然であると考えた.

3. 提案手法

以下,3 層のニューラルネットワークを MLP (Multiple Layer Perceptron)と表現する.本研究では複合的知識獲得モデルの問題点を解決するために,選択 MLP が試行錯誤しながら学習するのと同時に統合 MLP を並列学習させるモデルを提案する.逐次的な学習を廃止することによってモデルの学習を効率化し,より自然な知識獲得を実現することが期待できる.並列学習を行う際,選択 MLP は試行錯誤しながら学習を行うため,選択 MLP の選択した知識全てが正しいとは限らないという点を考慮する必要がある.そのため,並列学習を行う際,統合 MLP の学習に TD 誤差を用いることとする.TD 誤差は状態遷移の良し悪しを表す値であるため,誤った知識を教師信号として学習することを防ぐことが期待できる.学習方法として 2 種類の統合 MLP の学習方法を設定した.1 つ目は TD 誤差が正の値のときのみ学習を行う「TD 正学習」である.これにより,良い状態遷移を行った知識のみ教師信号として学習することが期待できる.2 つ目は,TD 誤差が正の値のときに加え,負の値の場合にも誤差逆伝播法の学習率を負の値に設定して学習させる「TD 正負学習」である.TD 正学習と比べ,悪い状態遷移を行った知識を行わないように学習することが期待できる.負の学習率として-0.005を用いることとする.

4. 計算機実験

本提案モデルの有効性を示すために「適応的障害物回避タスク」における統合知識獲得の計算機実験を行った.また,この実験ではシミュレータ webots ver7.2.0,シミュレーションには Khepera という移動型ロボットを用いた.このロボットは左右のタイヤによって移動し,距離センサ 8 つ,光センサ 8 つ,視覚センサを持つ.これらの各センサによって環境状態を観測する.タスクの環境を図 2 に示

Introduction of Parallel Learning to Combined Knowledge Acquisition Model
†Kohei Henmi, University of Yamanashi
‡Motonobu Hattori, University of Yamanashi

す.また,実験条件を表 1 に示す.

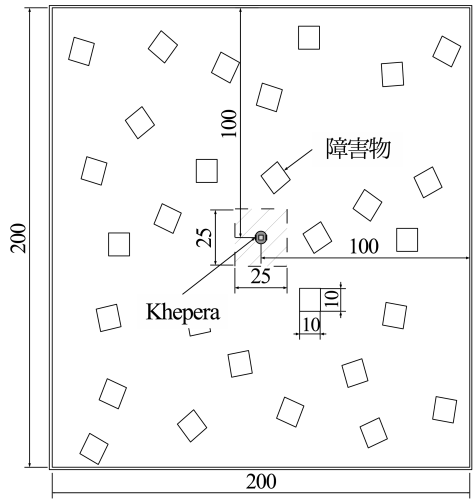


図 2: 適応的障害物回避タスク環境(単位:cm)

表 1: 計算機実験条件

学習率	0.1 / -0.005(TD 誤差正負のみ)
小報酬/罰	0.1 / 0.9
小報酬条件	直進を行う 障害物が右にある時,左に回避 障害物が左にある時,右に回避
試行成功条件	2000step 経過
試行失敗条件	障害物・壁に衝突(罰)

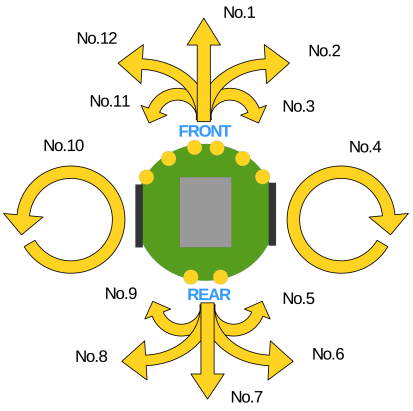


図 3: 12 個の既存知識(行動)

このタスクでは,環境内の障害物や壁を回避しながらできる限り直進を行うという知識を獲得することを目的としている.ロボットは初期位置固定,初期角度は毎試行 360 度ランダムとした.環境内には 10 ×10 ×10cm の障害物を 40 個配置し,各障害物間はロボットが通過できる距離を確保した.また, ロボットの初期位置の周囲に障害物配置禁止区域を設定した.今回のタスクでは 12 個の既存知識(図 3) に加え,「左旋回障害物回避」と「右旋回障害物回避」の 2 個の追加知識を与え,計 14 個の知識によって学習を行った.この追加知識はそれぞれ事前に同環境で学習して得た統合知識であり,左(右)に障害物を検知時,右(左)旋回を行うことができるようになっている.追加知識を加えることにより,適応的に左に障害物を検知したら右へ,右に障害物を検知したら左へそれぞれ回避行動をとる選択が学習できると考えられる.また,片側旋回のみによる回避やその場で回転することを防ぐため,テスト時の左右タイヤの移動変位の比から計算される直進率を設定した.この値が 1 に近い程適応的に回避を行えているということがわかる.今回は従来モデルと本提案モデルによる統合知識獲得までの学習試行回数を比較した.学習時には,10 試行毎に 100 回テストを行い,成功率 95%以上かつ直進率 0.9 以上となった場合,学習できたとみなした.各 10 シミュレーションを行ったときの平均学習試行回数を表 2 に示す.表では TD 誤差正学習を“TDp”,TD 誤差正負学習を“TDpn”と表している.

表 2: 平均学習試行回数

	従来モデル	TDp 学習	TDpn 学習
並列学習	–	239	181
選択規則学習	246	–	–
統合知識学習	365	–	–
学習試行回数	611±235.4	239±78.8	181±46.1

統合知識を獲得するまでの平均学習試行回数を比較すると,従来モデル−TDp 間,従来モデル−TDpn 間のそれぞれに有意水準 5%で有意差が確認された.提案モデルでは学習時の状態遷移の良し悪しを効率良く学習し,タスクを解くための十分な性能を持つネットワークを少ない試行回数で獲得できたといえる.また,TDp と TDpn では有意な差は確認されなかったが, TDpn の方が少ない試行回数で獲得できている.本タスクは,ある状態における適する知識と適さない知識が対極の関係にある(右側に障害物を検知したとき,適する知識が左旋回,適さない知識が右旋回)という性質を持つと考えられる.このことから,TD 誤差負の学習を加えることによってネットワークの学習が促進されたと考えられる.

5. まとめ

本研究では,複合的知識獲得モデルの効率化を目的とし,複合的知識獲得モデルを改良し,TD 誤差を用いて統合 MLP を並列学習させる方法を提案した.そして,性質の異なる複数のタスクにおいて計算機実験を行い,従来モデルに比べて知識獲得までの試行回数を大幅に削減することができた.また,逐次学習を廃止したことで,より自然な知識獲得ができたと考えられる.以上のことから,提案モデルの有効性が示された.

参考文献

[1] 矢部 達也,服部 元信:知識統合による複合的知識獲得モデルの特性および実環境への適応性の調査,情報処理学会全国大会講演論文集,Vol.70,No.2,pp.283-284(2008).

[2] 識名 翔,服部 元信:複合的タスクのための既存知識の選択規則の学習,情報処理学会全国大会講演論文集,Vol.72,No.2,pp.239-240(2010).