G-021

リカレントニューラルネットと高速統合学習を用いた Direct-Vision-Based 強化学習

Direct-Vision-Based Reinforcement Learning with Recurrent Neural Network Using Fast Integration Learning

井上 大輔 † 服部元信‡
Daisuke Inoue Motonobu Hattori

1 はじめに

従来よりロボットの自律学習の研究に強化学習が注目を集めてきた. 強化学習は報酬や罰といった少ない情報から行動を獲得するため非常に柔軟で自律性の高い学習が行える.

また近年, 強化学習に用いるアーキテクチャとしてニューラルネット (NN:Neural Network) が注目を集めている.NN とは, 生物の神経回路を模したネットワークモデルであり, 入力に対する識別能力が高い, 学習機能を有しているという特徴を持つ.

NN を用いた強化学習手法として, 柴田らによって提案された Direct-Vision-Based 強化学習 [1] がある. この手法では, 入力から出力までのシステム全体を NN によりシームレスに構成しており, ロボットに取り付けた視覚センサの値を直接 NN の入力に流し込む. そして, 視覚情報から状態評価を行い, 強化学習により教師信号を内部生成し NN を学習させる. そのため, 学習成果が他タスクに応用しやすい, 自律性が高く, システム全体を合目的的に学習できるといった特徴がある.

しかし, 学習時間が長い, 比較的容易なタスクへの適用しか確認されていないという問題がある. 学習時間に関しては, 強化学習部分での遅延と NN 学習部分での遅延の両方が考えられる. この手法では,NN の学習に誤差逆伝搬学習法 (BP:Back Propagation) を用いているが,BP には学習時間が長いという欠点がある. そのため,NN の学習手法の改良により学習時間の問題は軽減されると考えられる. また, より複雑なタスクの 1 つとして時系列学習を要する問題が挙げられる.

そこで本研究では, 学習の高速化, 時系列学習を要するタスクへの適用を目的とし, リカレントニューラルネット (RNN:Recurrent NN) と高速統合学習 (FIL:Fast Integration Learning)[2] を用いた強化学習手法を提案する.

2 高速統合学習 FIL

ここでは BP の高速化手法である FIL について説明する. FIL では, 次の評価関数 $J_1 \sim J_3$ に最急降下法を順次適用することで重みの学習を行う.

$$J_1 = -\sum_i \{t_i \ln o_i + (1-t_i)\ln(1-o_i)\} \quad (1)$$

$$J_2 = J_1 + \beta \sum_j |h_j - 0.5| + \varepsilon \sum_{i,j} |w_{ij}| \quad (2)$$

$$J_3 = J_1 + \beta \sum_{|h_j-0.5|<\theta_\beta} |h_j - 0.5| + \varepsilon \sum_{|w_{ij}|<\theta_\varepsilon} |w_{ij}| \quad (3)$$

ここで,o_i は出力ニューロン i の出力値,t_i はその教師信号, h_j は中間ニューロン j の出力値,w_{ij} はニューロン ij 間の結合重み, β, ε は各付加項の相対的重み,$\theta_\beta, \theta_\varepsilon$ は各付加項を選択的に適用させるための閾値である.

式 (1) の J_1 が FIL での誤差評価式となる. BP の重み修正における減衰項を除去した形となっており, 教師信号と出力との誤差のみに基づいた重みの修正を行うことで学習を高速化する. FIL では, まず J_1 を用い学習を高速に成功させ, 次に J_2, J_3 により構造学習を行う.

式 (2) の J_2 の右辺第 2 項を統合項, 第 3 項を忘却項と呼ぶ. 統合項では, 中間ニューロンへの出力値が 0.5 に近いものほど 0.5 に近づくように学習を進める. 忘却項では結合重みを全体的に 0 に近づける働きを持つ. これによりネットワークの規模を最適化するとともに, 学習に関与している部分を顕著にし骨格構造を得る. 骨格構造が得られると過学習を防ぐことができ, また, 類似する入力に対する反応が統一されるようになるので汎化性が向上するといえる.

最後に J_3 について説明する. J_2 で行った構造学習は,J_1 本来の学習を劣化させてしまう恐れがある. そこで,J_3 では, 閾値 $\theta_\beta, \theta_\varepsilon$ を用いて, 構造学習を適用する範囲を必要最小限にとどめることで, 望ましい学習到達度を得る.

3 提案手法

本研究では NN 構造とその学習法に着目した Direct-Vision-Based 強化学習の改良手法の提案を行う. また, 後述する二輪移動ロボットへの実装を考慮した設計となっている.

図 1 に示すのが従来法で用いるフィードフォワード型ニューラルネット (FFNN:Feed forward NN) である. 入力層に視覚センサ値が直接与えられ, 中間層を経て, 出力層では左右の車輪速度と状態評価値の 3 つの値が出力される. また, 提案手法では, 図 2 に示すようなリカレント構造を有しており,1 時刻前の中間層の値がそのまま文脈層として入力に与えられるほかに, 数時刻前までの車輪速度の出力値がバッファとして同じく入力に与えられる. これにより, 時系列データの学習が可能となる.

†山梨大学大学院医学工学総合教育部
‡山梨大学大学院医学工学総合研究部

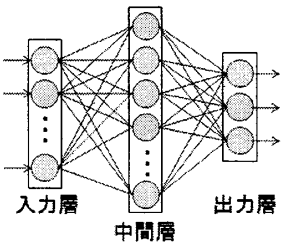

図 1: FFNN の構造.

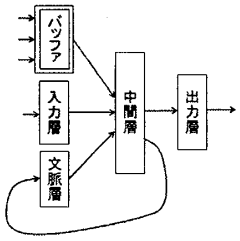

図 2: RNN の構造.

4 計算機シミュレーション

本手法の有効性を検証するため, ロボットシミュレータ Webots を用い, 二輪移動ロボット Khepera の制御に提案モデルを実装し, 計算機シミュレーションを行った. シミュレーションタスクとして, 視界の中に目標があればそこへ向かっていく「目標物到達」, 目標を視界の中心に捉えようとする「目標物探索」, これら二つを合わせた「目標物探索到達」, 地面に描かれたラインの上を進む「ライン追跡」を行った.

また, 実験に用いた NN 構造は, 入力層が 64 個, 中間層が 30 個, 出力層が 3 個のニューロンからなり, 加えて RNN では, バッファが 8 個, 文脈層が 30 個のニューロンからなる. そして, 一定試行ごとに 100 回のテストを行い, その成功率が 100%となれば学習完了とした.

表 1 に従来法との学習速度の比較結果を示す. 数値は学習完了までに要した試行回数の 20 回平均である. 結果より学習回数が遅いという従来法の問題点の改善ができたといえる. また, 提案法では構造化学習の結果として, 顕著な重みだけからなる骨格構造を得ることが可能であった. 具体的には, 重みの絶対値が 0.01 より小さい結合を不要な重みとした場合, ロボットの行動制御に悪影響を与えることなく, 全体の 97.1%の結合を削除することができた. 一方, 同様の処理を従来法に適用した場合,9.5%の結合しか削除されなかった. このことから,BP では, ほとんどの結合に重みを分散させて知識を形成しているのに対し,FIL では, 一部の結合のみに重みを分散させた非常にコンパクトな知識構造を得られることがわかる.

さらに, 探索到達タスクで学習時にはなかった障害物をロボットと目標物との間に設置したところ,BP で獲得した知識では障害物に関係なく目標に向かい障害物に衝突するのに対し,FIL で獲得した知識では障害物を回避するような動作を見せた (図 3,4). この差異は構造化の違いによるものと考えられる. 探索到達のタスクでは学習時に衝突による罰を与えられる機会は少ない. そのため,BP では分散された知識に埋もれてしまい表面化しなかったと考えられる. 一方の FIL では, 少ない学習であっても, ある程度の重みを持つと, それ以降の構造化の影響を受けにくくなり, 他の冗長なニューロンとの区別が顕著に長く残る. その差がこうして障害物に対する行動の差として表れたと考えられる.

ライン追跡タスクでは, ロボットは図 5 に示すラインに沿って移動する. このとき, 分岐点において観測される入力状態は同じだが, 右回り左回りで正解となる進路が異なる. そのため,1 時刻の入力しか与えられないと, ロボットは二つの状態を識別できず, 正しい右左折を行えない. そこで, 入力の他に過去の状態と出力を与えることで, その差異から二つの状態を識別可能とし, 常に正しい右左折が行える知識の獲得を目指した.

表 1: 学習回数の比較.

	探索	到達	探索到達
提案法 (RNN+FIL)	2613	970	2455
従来法 (FFNN+BP)	3080	2355	3635

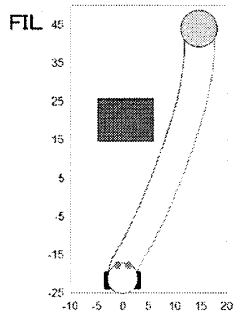

図 3: 提案法の獲得知識.

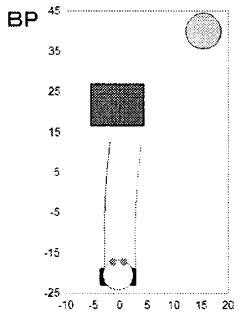

図 4: 従来法の獲得知識.

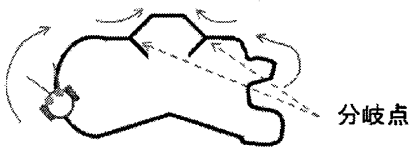

図 5: ライン追跡タスク.

その結果, 間違ったルートを選択した場合, 当初はそのまま直進してラインを見失い試行失敗となっていたが, 学習が進むにつれ, 途中で停止, 後退し, 正しい進路に復帰するという行動が確認できた. 結論として, 時系列を学習していると思われる動作が獲得できた.

5 まとめ

本研究では,Direct-Vision-Based 強化学習における学習の高速化, 時系列学習を要するタスクへの適用を目的とし,BP の高速化手法である FIL と RNN による時系列の学習機構を取り入れたモデルを提案した. 結果として学習の高速化に成功し, 時系列データの学習も行うことができた. また, 構造学習による汎化性能の向上を確認できた.

参考文献

[1] 柴田克成, 岡部洋一, 伊藤宏司,“ニューラルネットワークを用いた Direct-Vision-Based 強化学習 センサからモータまで,” 計測自動制御学会論文集, Vol.37, No.2, pp.168-177, 2001.

[2] 菊地進一, 中西正和,“短期記憶を用いたリカレントニューラルネットワークと高速な構造学習法,” 電子情報通信学会論文誌, Vol.J84-D-2, No.1, pp.159-169, 2001.