RTRL を用いた強化学習による予測機能の自律的獲得

蔡　詩祐†　　平原　誠†

† 法政大学大学院理工学研究科

1. はじめに

階層型リカレントニューラルネットを用いた強化学習アルゴリズムが提案されている[2]．ニューラルネットの学習則には，BPTT(Back Propagation Through Time)が用いられていた．予測機能なしでは達成が困難なタスクに対し，学習者(エージェント)に予測機能を自律的に獲得させている．この手法の有効性の検証に用いられたタスクは，開始状態と終了状態をエージェントに明示できるもの，即ちエピソード的タスクであった．しかし現実世界では，開始状態と終了状態を具体的に明示できない問題(連続的タスク)が多く存在する．本研究では，先行研究で扱われた階層型リカレントニューラルネットを全結合型リカレントニューラルネットへ変更し，学習則には RTRL(Real Time Recurrent Learning)[3]を用いる．これにより実時間での学習を可能にし，連続的タスクへのアプローチを図る．

2. 提案手法

本研究では入力層を除き，すべての素子にフィードバック結合をもつ全結合型リカレントニューラルネットを用いる．時刻tにおけるタスクの環境を表す状態ベクトルをs_tとすると，入力層出力ベクトル$x(t)$は，

$$x(t) = s_t \tag{1}$$

と表せる．時刻tにおける入力層素子iの出力$x_i(t)$，中間層，出力層素子iの出力$y_i(t)$をまとめて，

$$z_i(t) = \begin{cases} x_i(t), i \in I \\ y_i(t), i \in H \cup O \end{cases} \tag{2}$$

と表すことにする．ここでI, H, Oはそれぞれ入力，中間，出力層素子の集合である．また$y_i(t)$は，

$$y_i(t) = f\big(net_i(t)\big), i \in H \cup O \tag{3}$$

$$f(net_i(t)) = \frac{1}{1 + e^{-net_i(t)}} - 0.5 \tag{4}$$

$$net_i(t) = \sum_{j=I \cup H \cup O} w_{ij} z_j(t-1) \tag{5}$$

と表せる．$net_i(t)$は時刻tでの素子iの内部値，w_{ij}は素子jからiへの結合重みである．出力層素子の出力を各行動の Q 値として扱い，

$$Q_i(t) = y_i(t), i \in O \tag{6}$$

と表す．時刻tで選択した行動a_tの行動価値$Q_{a_t}(t)$への教師信号$T_{a_t,t}$は，

$$T_{a_t,t} = r_{t+1} + \gamma \max_{a'} Q_{a'}(t+1) \tag{7}$$

で与えられる．ここでγは 0 以上 1 未満の係数，rは報酬である．時刻tでのw_{ij}の更新は，μを学習係数，t_0を初期時刻とすると，

$$\Delta w_{ij}(t) = \mu \sum_{k \in H \cup O} e_k(t) p_{ij}^k(t) \tag{8}$$

$$e_i(t) = \begin{cases} T_{a_t,t} - Q_i(t) & (if\ i = a_t) \\ 0 & (otherwise) \end{cases}, i \in H \cup O \tag{9}$$

$$p_{ij}^k(t) = f'\big(net_k(t)\big)\left[\sum_{l \in H \cup O} w_{kl} p_{ij}^l(t-1) + \delta_{ik} z_j(t-1)\right] \tag{10}$$

$$\delta_{ik} = \begin{cases} 1 & (if\ i = k) \\ 0 & (otherwise) \end{cases} \tag{11}$$

$$p_{ij}^k(t_0) = 0 \tag{12}$$

この手法の場合，時刻$t+1$での行動価値$Q(t+1)$の値は，時刻tでの素子出力$z_i(t)$から計算できるので，時刻$t+1$での素子出力に影響を受けない．したがって，時刻tの時点で式(7)より時刻tの教師信号$T_{a_t,t}$を求めることができ，ひいては RTRL による実時間での学習が可能となる．時系列展開した提案手法のニューラルネットを図 1 に示す．

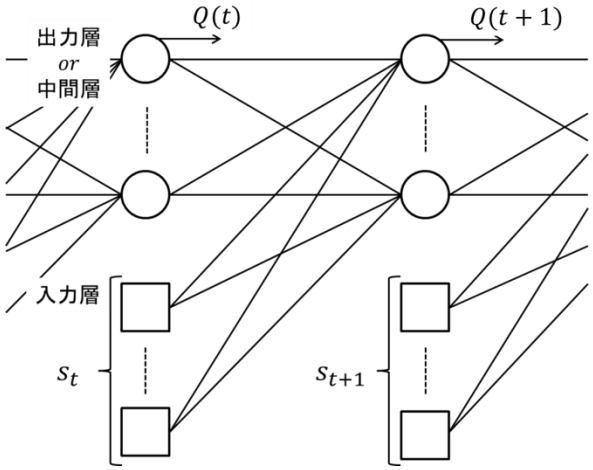

図 1.　時系列展開した提案手法のニューラルネット

参考文献

[1] R.S.Sutton and A.G.Barto: Reinforcement Learning, MIT Press, 1998.
[2] Kenta Goto and Katunari Shibata: Eemergence of Prediction by Reinforcement Learning Using a Recurrent Neural Network, Journal of Robotics, Volume 2010, Article ID437654.
[3] Ronald J. Williams and David Zipser: A Learning Algorithm for Continually Running Fully Recurrent Neural Networks, Neural Computation, 1, pp.270-280, 1989.