修士論文 要旨

スパイキングニューラルネットワークと報酬調節シナプス可塑性に基づく強化学習

数理情報学専攻 48216222　鶴海 杭之
指導教員　田中 剛平 特任准教授

1　はじめに

機械学習の分野において，人工ニューラルネットワーク（ANN）は欠かせない技術となっている．しかし，大量のパラメータを学習するのに十分な大量の教師データの取得や，学習に要する計算コストの増加，それに伴う電力消費量などは課題となっている [4]．教師データとなる，モデルが出力すべき目標となるデータがない場合には，教師データを使用しない教師なし学習や，モデル出力の良し悪しを示す間接的な情報によって学習する強化学習が用いられる．本研究ではそのうちの強化学習に着目する．

強化学習は，動的計画法に端を発し TD 学習 [3] を中心として数理的側面の強い手法が手法が続々と開発されてきた一方で，神経科学の文脈で，人間を含む動物の学習との関係も注目されている．その中で，シナプス可塑性をもとにした強化学習の学習則が研究されている [1]．本研究ではそのような学習則を用いた強化学習にフォーカスする．

一方，スパイキングニューラルネットワーク（SNN）は，ANN に比べて生物学的妥当性があり，発火イベントで情報を表現できるため，ハードウェア実装によって計算時間や電力消費量を抑えた情報処理につながると期待されている．SNN とシナプス可塑性に基づく学習則で強化学習を行った研究はまだ数が多いとは言えず，どのような要因が性能に影響するかは未知な部分が多い．本研究では，そのような強化学習において結合の仕方や学習則の適用の仕方が学習性能にどう影響するかを調べる．

2　SNN モデル

2.1　モデルの構造

実験で使用するモデルは，入力層，中間層，出力層の 3 層からなる．

入力層はタスクに合わせた設計になっている．中間層内の結合は，自己結合なしで，それ以外は確率 0.1 で結合している．入力層から中間層への結合は全結合とする．出力層はニューロン 1 個である．中間層から出力層への結合は全結合である．なお，中間層内の結合強度を 0 にするとフォードフォワードニューラルネットワーク（FNN）の構造になる．

ニューロンモデルには次の式で記述される LIF モデル [2] を用いる．

$$\tau_m \frac{\mathrm{d}v}{\mathrm{d}t} = -(v - v_r) + RI, \tag{1}$$

$$\text{if } v \geq \theta, \text{ then } v = v_0. \tag{2}$$

ここで v はニューロンの膜電位，τ_m は膜時定数，v_r は静止膜電位，R は膜抵抗，I はニューロンへの入力電流，θ は閾値，v_0 はリセット電位を表す．$\tau_m = 20\,\mathrm{ms}$，$v_r = -70\,\mathrm{mV}$，$R = 1$，$\theta = -54\,\mathrm{mV}$，$v_0 = -70\,\mathrm{mV}$ とする．

2.2　シナプス伝達

シナプス伝達は次のようにモデル化する．

$$v_i \leftarrow \begin{cases} v_i + \alpha_{ij} w_{ij} & \text{（興奮性ニューロン } j \text{ 発火時）} \\ v_i - \alpha_{ij} w_{ij} & \text{（抑制性ニューロン } j \text{ 発火時）} \end{cases}, \tag{3}$$

ここで v_i はニューロン i の膜電位を，w_{ij} はニューロン j からニューロン i へのシナプス結合の重みを表す．重み w_{ij} の初期値は $[0,\ w_{\max}]$ 上の一様分布から定める．可塑性がある場合 w_{ij} は変動するが，$[0, w_{\max}]$ の範囲に収まるようにクリッピングする．$w_{\max} = 5\,\mathrm{mV}$ とする．α_{ij} はモデルの挙動を調整するためのパラメータで，ニューロン i, j がどの層に属するかによって値を決める．α_{ij} を変えることは $w_{\max}$ を変えるのと同等の効果を持つ．シナプス伝達の遅延を 1 秒とする．

2.3　シナプス可塑性

結合重みの変化が報酬に依存するような学習則である reward-modulated spike-timing-dependent plasticity（RM-STDP）[1] を用いる．RM-STDP は

$$\frac{\mathrm{d}w_{ij}(t)}{\mathrm{d}t} = \gamma r(t) z_{ij}(t) \tag{4}$$

と記述される．ここで γ は学習率，$r(t)$ は報酬であり，$z_{ij}(t)$ は eligibility trace と呼ばれる変数で，シナプス前細胞 j とシナプス後細胞 i の発火の履歴を保持している．$z_{ij}(t)$ は次の式に従う．

$$\tau_z \frac{\mathrm{d}z_{ij}(t)}{\mathrm{d}t} = -z_{ij}(t) + \xi_{ij}(t). \tag{5}$$

ここで τ_z は eligibility trace の時定数で，今回の実験では $\tau_z = 100\,\mathrm{ms}$ とする．$\xi_{ij}(t)$ は次の式で定義される．

$$\xi_{ij}(t) = P_{ij}^{+}(t)\Phi_i(t) + P_{ij}^{-}(t)\Phi_j(t), \tag{6}$$

$$\frac{\mathrm{d}P_{ij}^{+}(t)}{\mathrm{d}t} = -\frac{P_{ij}^{+}(t)}{\tau_+} + A_+\Phi_j(t), \tag{7}$$

$$\frac{\mathrm{d}P_{ij}^{-}(t)}{\mathrm{d}t} = -\frac{P_{ij}^{-}(t)}{\tau_-} + A_-\Phi_i(t). \tag{8}$$

Φ_i はディラックのデルタ関数及びそれを平行移動したものを足し合わせたもので，ニューロン i が発火した時刻のみ値が 0 でないようなものとする．A_+，τ_+，τ_- は正の定数，A_- は負の定数であり，今回の実験では $A_+ = 0.05$，$A_- = -0.05$，$\tau_+ = \tau_- = 20\,\mathrm{ms}$ とする．

本実験において，中間層から出力層への結合には常に RM-STDP を適用する．入力層から中間層への結合については RM-STDP を適用する条件と可塑性がない条件の両方で実験を行う．

3 数値実験

3.1 タスク

排他的論理和（XOR）問題と ON/OFF 切り替えのタスクの，2 つのタスクで実験を行う．XOR 問題はもともと教師あり学習の問題だが，ここでは RM-STDP を提唱した論文 [1] と同様に，報酬を導入して強化学習のタスクとして扱う．

XOR 問題では，入出力を入力層ニューロン・出力層ニューロンの発火率の高低で表現し，正しい出力が 1 のときに出力層ニューロンの発火率が上がり 0 のときに下がるように報酬を与える．(0, 0), (0, 1), (1, 0), (1, 1) の 4 通りの入力をランダムな順番で 500 ms ずつ提示し 1 エポックとする．1 回の実験でこれを 100 エポック行う．

ON/OFF 切り替えのタスクでは，入力層ニューロンは 2 つで，それぞれの発火が ON, OFF を示す信号である．ON を示す信号が来たら出力ニューロンの発火率が高くなり，OFF を示す信号が来たら出力ニューロンの発火率が低くなることを目標とする．1 回の実験で 50 秒のシミュレーションを行う．

3.2 結果

XOR 問題の実験において，中間層内部に結合がない場合の報酬の推移は図 1 のようになった．入力層から中間層への結合に可塑性がないと学習に失敗し，可塑性があると学習に成功したという結果である．中間層内部に結合がある場合も同様の結果となった．より詳細な解析を行ったが，中間層内部の結合が学習に与える影響ははっきりしなかった．

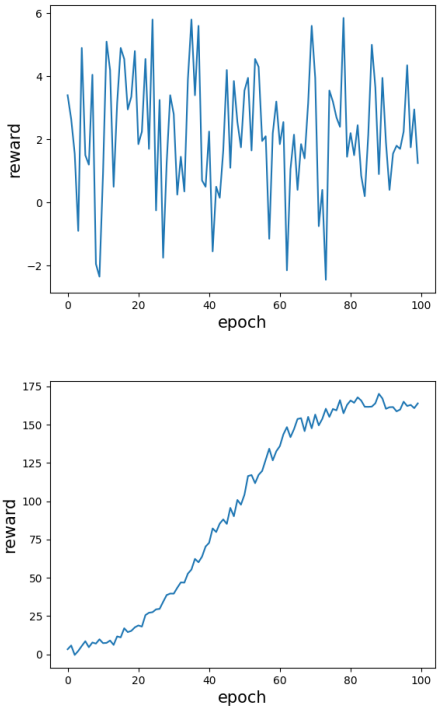

図 1. 中間層内部に結合がない場合で，入力層から中間層への結合に可塑性がない条件（上），ある条件（下）での報酬の推移（20 回の平均）

ON/OFF 切り替えのタスクにおいては，中間層内部に結合がない場合とある場合のいずれにおいても，入力層から中間層への結合に可塑性があってもなくても学習に成功した．より詳細な解析を行い，中間層内部の結合は学習に悪影響を与えていると推測された．

これらの結果から，入力層から中間層への結合の可塑性はあった方が学習に有利であること，中間層内部の結合はタスクによっては学習に悪影響を与えることが示唆された．

参考文献

[1] Răzvan V Florian. Reinforcement learning through modulation of spike-timing-dependent synaptic plasticity. *Neural computation*, Vol. 19, No. 6, pp. 1468–1502, 2007.

[2] Wulfram Gerstner and Werner M Kistler. *Spiking neuron models: Single neurons, populations, plasticity*. Cambridge university press, 2002.

[3] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.

[4] Vivienne Sze, Yu-Hsin Chen, Tien-Ju Yang, and Joel S Emer. Efficient processing of deep neural networks: A tutorial and survey. *Proceedings of the IEEE*, Vol. 105, No. 12, pp. 2295–2329, 2017.