

ニューロンの追加及び重みベクトルの分布に基づいた想起が可能な KFM 連想メモリによる拡張適格度を用いた強化学習

佐藤宏樹 長名優子

東京工科大学 コンピュータサイエンス学部

1 はじめに

環境との相互作用により適切な行動系列を獲得するための学習手法として、強化学習に関する様々な研究が行われている [1].

本研究では、ニューロンの追加及び重みベクトルの分布に基づいた想起が可能な KFM (Kohonen Feature Map) 連想メモリを提案し、それを用いた拡張適格度を用いた強化学習を実現する。提案モデルでは、学習時に学習データと関連づけられた領域として適切なサイズの領域を確保できない場合に、すでに学習済みの領域のサイズを縮小し、新たにニューロンを追加することで、さらに学習を行えるようにしている。

2 ニューロンの追加及び重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリ

2.1 構造

提案モデルは、入出力層とマップ層から構成されており、入出力層は複数のパターンを表す部分に分けられている。

2.2 学習過程

提案モデルでは、学習が十分に進んだ重みを固定することで既学習パターンの記憶を破壊することなく新しいパターンを学習することができるようにしている。また、学習パターンをマップ層の重みが固定されたニューロンを中心とする楕円形の領域に対応させることで学習を行うが、その際に近傍領域も含めて学習を行うことで、類似したパターンに対応する領域がマップ層上の近い位置に配置されるようにしている。提案モデルの逐次学習アルゴリズムを以下に示す。

- (1) 重みの初期値をランダムに選ぶ。

Reinforcement Learning by Variable-Sized Kohonen Feature Map Probabilistic Associative Memory using Extended Eligibility
Hiroki Sato and Yuko Osana (Tokyo University of Technology, osana@cs.teu.ac.jp)

- (2) 学習ベクトル $X^{(p)}$ と重みベクトル W_i のユークリッド距離 $d(X^{(p)}, W_i)$ を計算する。

- (3) マップ層のすべてのニューロンに対して $d(X^{(p)}, W_i) > \theta^t$ のとき、入力されたパターン $X^{(p)}$ は未学習であると判断され、学習を行う。既学習であると判断された場合には (13) に進む。

- (4) 重みが固定されたニューロンが1つも存在しない場合は、学習ベクトルとマップ層のニューロンの重みベクトルが最小となるニューロンを学習する領域の中心 c として決定し、(12) に進む。1つでも固定されたニューロンが存在する場合には、(5) に進む。

- (5) 既学習パターンに対応する領域と重複せずに入力されたパターン $X^{(p)}$ を学習するための領域を確保することができるかを調べる。学習する領域の候補となるニューロンが存在する場合には (12) に進む。候補となるニューロンが1つも存在しない場合には、(6) に進む。

- (6) 既学習パターンに対応する領域に含まれるニューロンのニューロン間の距離を $\phi_n(d_z^{min})$ に縮めることで入力されたパターン $X^{(p)}$ を学習するための領域を確保することができるかを調べる。ここで、 $\phi_n(\cdot)$ は

$$\phi_n(d) = \begin{cases} d/2^n, & d/2^n > d^{min} \text{ のとき} \\ d & \text{それ以外} \end{cases} \quad (1)$$

で与えられる。ここで、 n はパターン $X^{(p)}$ が入力されてから (5) において領域を確保できるかを調べた回数、 d_z^{min} は固定ニューロン z から最も近くにあるニューロンとの距離、 d^{min} はマップ層における隣接するニューロンとの距離の取り得る最小値である。学習する領域の候補となるニューロンが存在する場合には (9) に進む。候補となるニューロンが1つも存在しない場合には、(7) に進む。

- (7) 既学習パターンに対応する領域に含まれるニューロンのニューロン間の距離を $\phi_n(d_z^{min})$ に縮め、

入力されたパターン $X^{(p)}$ を学習するための領域を領域に含まれるニューロンのニューロン間の距離が $\phi_n(d^{max})$ となるようにした場合に、領域が確保できるかを調べる。ここで、 d^{max} はマップ層における隣接するニューロンとの距離の取り得る最大値であり、提案モデルでは $d^{max}=1$ としている。学習する領域の候補となるニューロンが存在する場合には (9) に進む。候補となるニューロンが 1 つも存在しない場合には、(8) に進む。

- (8) (6), (7) における n に対し、更に領域のサイズを縮小できれば (6) に戻る。そうでない場合には、これ以上新しいパターンを学習するための領域を確保できないものとして学習を終了する。
- (9) 学習する領域の中心の候補として選ばれたニューロンの中から学習ベクトルと重みベクトルが最小となるニューロンを学習する領域の中心となるニューロン c として選択する。
- (10) (6) または (7) で学習する領域の中心の候補が選ばれた場合には、既学習パターンに対応する領域に含まれるニューロン間の距離を必要に応じて縮小し、ニューロンの追加を行う。
- (11) (6) または (7) で学習する領域の中心の候補が選ばれた場合には、新しく学習するパターン $X^{(p)}$ に対応する領域に該当する部分に必要なに応じてニューロンの追加を行う。
- (12) $d(X^{(p)}, W_r) > \theta^t$ のとき、重みが固定されていないニューロンに結合する重みを更新する。
- (13) ニューロン c に結合する重み W_c を固定する。
- (14) 新しいパターンが入力されるたびに (2)~(13) を繰り返す。

2.3 想起過程

提案モデルにおいて、パターン X が入出力層に与えられた場合には、マップ層のニューロン i の出力 x_i^{map} は

$$x_i^{map} = \begin{cases} 1, & i = r \text{ のとき} \\ 0, & \text{それ以外} \end{cases} \quad (2)$$

で与えられる。なお、 r は勝ちニューロンであり、

$$\frac{1}{N^{in}} \sum_{k \in C} g(X_k - W_{ik}) \geq \theta^{map} \quad (3)$$

を満たすニューロンの中からランダムに 1 つ選ばれる。ここで、 N^{in} は入力を与えられた入出力層のニューロン数、 C は入出力層において入力を与えられたニューロンの集合である。また、 $g(\cdot)$ は

$$g(h) = \begin{cases} 1, & |h| < \theta^d \text{ のとき} \\ 0, & \text{それ以外} \end{cases} \quad (4)$$

で与えられる。ここで、 θ^d はしきい値を表す。また、式 (3) において θ^{map} はマップ層ニューロンのしきい値である。

バイナリパターン X が入出力層に与えられた場合、入出力層のニューロン k の出力 x_k^{io} は以下のように計算される。

$$x_k^{io} = \begin{cases} 1, & W_{rk} \geq 0.5 \text{ のとき} \\ 0, & \text{それ以外} \end{cases} \quad (5)$$

アナログパターン X が入出力層に与えられた場合、入出力層のニューロン k の出力 x_k^{io} は以下のように計算される。

$$x_k^{io} = W_{rk} \quad (6)$$

3 拡張適格度を用いた強化学習の実現

本研究では、2 で述べたニューロンの追加及び重みベクトルの分布に基づいた想起が可能な KFM 連想メモリをアクター・クリティック [2] のアクターの部分に用いることで強化学習を実現する。文献 [3] の研究では重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリ [4] を用いて強化学習を実現しているが、ここでは同様の手法をニューロンの追加及び重みベクトルの分布に基づいた想起が可能な KFM 連想メモリを用いて実現する。

4 計算機実験

提案するニューロンの追加及び重みベクトルの分布に基づいた想起が可能な KFM 連想メモリにおいて、1 対多の関係にあるパターンの逐次的な学習が行えること、必要に応じてマップ層のニューロンの追加が行えることを確認した。また、ニューロンの追加及び重みベクトルの分布に基づいた想起が可能な KFM 連想メモリによる拡張適格度を用いた強化学習を獲物捕獲問題に適用し、獲物を効率的に捕獲できるような行動を学習できることを確認した。

参考文献

- [1] R. S. Sutton and A. G. Barto : Reinforcement Learning, An Introduction, The MIT Press, 1998.
- [2] I. H. Witten : "An adaptive optimal controller for discrete-time Markov environments," Information and Control, Vol.34, pp. 286-295, 1977.
- [3] K. Watanabe and Y. Osana : "Reinforcement learning by improved Kohonen feature map probabilistic associative memory based on weights distribution using extended eligibility," SMC, Anchorage, 2011.
- [4] S. Noguchi and Y. Osana : "Improved Kohonen feature map probabilistic associative memory based on weights distribution," IJCNN, Barcelona, 2010.