

重みベクトルの分布に基づいた想起が可能な 改良型 KFM 連想メモリによる強化学習の実現

野口伸吾 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。強化学習では、設計者が意図しない未知の環境やノイズの多い実環境においても学習が行えるという特徴がある [1][2]。

本研究では、重みベクトルの分布に基づいた想起が可能な改良型 KFM(Kohonen Feature Map) 連想メモリを提案する。このモデルは、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 [3] に基づいたモデルであり、重みの更新方法を変更することで学習の高速化を実現している。また、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をマップ層上の近い位置に配置できるようにしている。さらにこのモデルを強化学習に導入し、重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリによる強化学習を実現する。提案手法では、試行中に選択された行動とそのときの状態の対に対応する領域のサイズを報酬に応じて変更するだけでなく、適格度に基づいて訪問頻度の少ない状態に関する領域のサイズを縮小することで重みベクトルの分布に反映されるようにしている。

2 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

ここでは、提案する重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリについて説明する。提案モデルは、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 [3] に基づいたモデルであり、重みの更新方法を変更することで学習の高速化を実現している。また、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をマップ層上の近い位置に配置できるようにしている。

Reinforcement Learning by Improved KFM Probabilistic Associative Memory based on Weights Distribution
Shingo Noguchi and Yuko Osana (Tokyo University of Technology, osana@cs.teu.ac.jp)

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、重みベクトルの分布に基づいた想起が可能な KFM 連想メモリと同様に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。また、重みの更新方法を変更することで学習の高速化を実現している。

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。
- (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 (7) に進む。
- (4) 学習する領域の中心となるニューロン r を以下のように決定する。

$$r = \underset{i: D_{iz} + D_{zi} < d_{iz}}{\operatorname{argmin}} d(X^{(p)}, W_i) \quad (1)$$

(for $\forall z \in F$)

ここで、 D_{ij} はニューロン i を中心とする領域のニューロン j の方向への動径、 d_{iz} はニューロン i と重みが固定されているニューロン z との距離、 F は重みが固定されたニューロンの集合を表す。なお、提案モデルではマップ層を 2 次元のトラスとして扱う。式 (1) により、新たに入力され

たパターンに対応する領域を既学習パターンに対応する領域と重ならないように確保できるようなニューロンの中で学習ベクトルに最も類似した重みベクトルを持つニューロンが学習する領域の中心として選択される。

- (5) $d(\mathbf{X}^{(p)}, \mathbf{W}_r) > \theta^t$ 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以下のように更新する。

$$\mathbf{W}_i(t+1) = \begin{cases} \mathbf{X}^{(p)}, & \theta_1^{\text{learn}} \leq H(\overline{d_{ri}}) \text{ のとき} \\ \mathbf{W}_i(t) + H(\overline{d_{ri}})(\mathbf{X}^{(p)} - \mathbf{W}_i(t)), & \theta_2^{\text{learn}} \leq H(\overline{d_{ri}}) < \theta_1^{\text{learn}} \text{ かつ} \\ H(\overline{d_{i^*i}}) < \theta_1^{\text{learn}} \text{ のとき} \\ \mathbf{W}_i(t), & \text{それ以外} \end{cases} \quad (2)$$

ここで、 $\theta_1^{\text{learn}}, \theta_2^{\text{learn}}$ はしきい値である。また、 $H(\overline{d_{ri}}), H(\overline{d_{i^*i}})$ は式 (3) のような準固定を実現するための関数であり、 $H(\overline{d_{ri}})$ は近傍関数の役割も果たしている。なお、ここで i^* はニューロン i から最も近い位置にある重みが固定されているニューロンを表す。

$$H(\overline{d_{ij}}) = \frac{1}{1 + \exp\left(\frac{\overline{d_{ij}} - D}{\varepsilon}\right)} \quad (3)$$

ここで、 $\overline{d_{ij}}$ はニューロン i が領域の中心であるときのニューロン i とニューロン j の距離 d_{ij} をニューロン i を中心とする領域のニューロン j 方向への動径 D_{ij} で割ることで正規化したものである。また、式 (3) において D ($1 < D$) は近傍領域のサイズを決める定数、 ε は関数の傾きを決める係数である。なお、重みが固定されているニューロンが存在しない場合には、 $H(\overline{d_{i^*i}}) = 0$ とする。

- (6) ニューロン r に結合する重み $\mathbf{W}_r$ を固定する。
(7) 新しいパターンが入力されるたびに (2)~(6) を繰り返す。

2.3 想起過程

提案モデルは重みベクトルの分布に基づいた想起が可能な KFM 連想メモリと同様の方法で想起を行う。

3 重みベクトルの分布に基づいた想起が可能な KFM 連想メモリによる強化学習の実現

アクター・クリティックのアクターの部分を、2 で述べた重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリを用いて実現する。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

モリの入出力層を状態と行動を表す部分に分け、状態を入力としたときに行動を出力できるように学習を行う。本研究では、環境から観測された情報だけでなく、エージェント自身が直前にとった行動と直前に観測した環境も状態として用い、とるべき行動の判断を行う。また、(1) 重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリによって現在の環境と直前の行動と直前の環境をもとに選択した行動、(2) 重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリによって現在の環境のみをもとに選択した行動、(3) エージェントがとることのできる行動の中からランダムに選択した行動の 3 つの行動のうち、その行動をとったときに環境から受け取る報酬が大きくなるような行動を選択する。このような行動の選択を行うことで学習の初期段階においても、最初に学習した行動のみをとりつづけることなく、試行錯誤を行うことが可能となる。

クリティック部分では環境から得られる状態を入力とし、価値の更新や評価を行う。さらに、アクターへ TD 誤差を出力する。アクターとして動作する重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリでは、報酬に基づいて学習を行い、環境の状態から行動の選択を行う。

4 計算機実験

提案モデルの動作を確認し、有効性を示すために計算機実験を行った。提案モデルにおいて、1 対多の関係にあるパターンに対応する領域のサイズを変えて学習し、領域のサイズに応じた確率で共通項に対応するパターンが想起できることを確認した。また、マップ層ニューロンの破壊やノイズに対するロバスト性があることなどを確認した。また、経路探索問題を例題として重みベクトルの分布に基づいた想起が可能な改良型 KFM 連想メモリを用いた強化学習の動作を確認し、有効性を示すために計算機実験を行った。提案手法において、報酬に基づいて環境に応じた領域のサイズの修正が行えること、適格度に基づいて訪問頻度の少ない状態に関する領域のサイズの縮小が行えることなどを確認した。

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] M. Koike and Y. Osana : "Kohonen feature map probabilistic associative memory based on weights distribution," Proceedings of IASTED Artificial Intelligence and Applications, Innsbruck, 2010.