論 文 の 内 容 の 要 旨

論文題目　　強化学習のための自律分散型関数近似法

氏　名　　小 林 祐 一

従来の離散的・記号的表象を基礎とした強化学習研究では重視されてこなかったが，強化学習の実問題への適用においては，情報表現の効率が学習性能に大きな影響を与える．つまり，強化学習の核をなす価値関数の更新アルゴリズムにおいて適切な価値関数の関数近似を行うことが重要である．強化学習における価値関数近似には，一般に

- 非定常関数近似に適していること
- ブートストラップ型の関数近似において収束性がある程度保証されること
- 学習に必要なデータを最小限にとどめること

などの要件が存在する．分解能の高い関数近似を行うためには，関数近似要素を増やすため必要メモリが増加することに加え，学習効率の問題が存在する．分解能を高くするためにはより多くの学習データが必要であるが，学習に必要なデータの増大は行動の試行錯誤量の増大につながり学習性能の低下を招く．このような意味から，学習にとって重要な領域のみに高い分解能を持つような表現，すなわち**適応的な分解能**を獲得する関数近似手法が重要な意味を持っている．

適応的な分解能を自律的に獲得するための関連研究として，関数近似手法における逐次的な基底関数の追加方法や状態空間の自律的な分割手法などをあげることができるが，人手による調整が必要な設計要素が多いという問題が存在する．それに対し，本研究では自律的な情報表現獲得を目指し，

を提案した．強化学習における適応的に分解能を獲得する関数近似でこれまで注目されてきたのは近似精度や学習達成度であったが，本研究では価値関数の形状に着目し，勾配変化を適切に表現できるように勾配変化の大きい領域に密に，小さい領域に疎に分布するように適応的に関数近似要素を再配置するアルゴリズムを提案した．系 (関数近似器) 全体の秩序との関係を記述した形での局所的な関数近似要素の挙動を設計する手法としてグラフ上の反応拡散方程式を用いた方法を提案した．

強化学習適用で想定するのは多次元入力 1 次元出力の陽関数であり，本研究では位置と勾配をもったノードによる超平面の組み合わせによって関数近似面を構成する．このノードを近傍を表すアークで結ぶことでグラフを構成し，各ノード近傍における近似関数の複雑度を定義する (**図.1**)．各ノードの複雑度を均一にするようにノードを移動させる (**図.2**)．勾配変化を反映した複雑度を定義することにより，ノードが勾配変化の大きい領域に密に，小さい領域に疎に分布するように挙動を設計する．そのための設計方法としてグラフ上の反応拡散方程式を用いた．

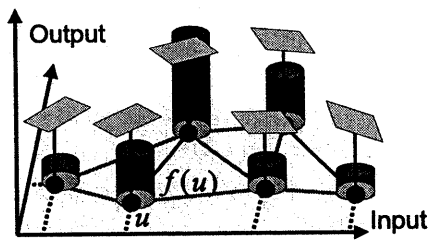

図 1: 複雑度の定義

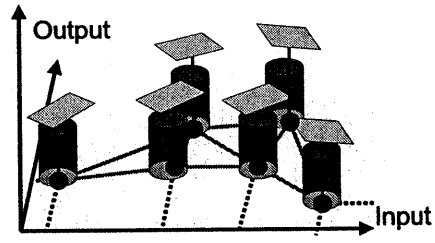

図 2: 複雑度の均一化

グラフ上の反応拡散方程式は自律分散システムの理論であり，不均一に分布するサブシステムからなるシステムにおいて，全体としての秩序を形成するためのサブシステムの挙動を勾配系に基づいて設計する手法である．提案手法は，自律分散系のアルゴリズムの特徴であるシステムの拡縮性に優れるという特質をもつため，ノードの動的な追加・削除が容易であるという利点を有する．位置と勾配情報を持ったノードの近傍をアークで結ぶことによって境界付きグラフに適用を行う．このとき，ノード間のアークを構成する方法として，ユークリッドノルムに基づいた近傍を構成する TRN(Topology Representing Networks) アルゴリズムを用いた．

複雑度の定義には，各ノードの近傍ノードに対する勾配変化と位置変化の積として定義した．この定義については，直観的には勾配変化とノードの密度が比例関係にあることと複雑度が一定になることが1次元では等価になることで設計指針に沿ったものであるという理解ができる．さらに，1次元においてSpline補間曲線の理論を用いて曲線近似誤差最小化問題との対応が取れることを示した．提案手法がノードの動的な追加を容易に行える方法であることに関して，また，近似目標関数を周波数領域に変換しサンプリング定理を適用することで，関数値のとるべき水準についての考察を行った．ノードの逐次的な追加に際しての関数近似がすでに充分な精度で達成されているかどうかの判断基準を従来研究のような近似誤差に基づく方法ではなく関数形状情報に基づく方法で目安を示した．

提案関数近似手法の強化学習への適用方法として，Value Gradient 法およびQ-learning 法の状態価値関数および行動価値関数の近似を行う方法を示した．Value Gradient 法では，状態価値関数をTD 誤差学習により推定し，行動決定は近似した状態価値関数の勾配情報を用いて行う．Q-learning 法では，要素行動の数だけ関数近似器を独立に用意し，各関数近似器は特定の要素行動に関する行動価値関数の近似を行う．プログラム実装に際しては，入力次元やノード数の増大に伴う計算量変化について考察を行った．

提案する関数近似手法の性質の検証および強化学習例題を用いた性能向上の確認を行った．定常関数近似問題においては，位置推定および勾配推定の両者についてノード移動を適応的に行うことにより近似誤差を低減できることを確認した．また，複雑度を表す関数値の大きい領域に逐次的にノードを追加していく方法が複雑な形状の関数を近似する上で有効に機能することを確認した．**図.3** は1次元，**図.4** は2次元のガウス関数を近似し，ノード移動により複雑な領域に高い分解能を得るようにノードの再配置が達成されていることがわかる．1次元ガウス関数の近似問題では，同数のノードに対し，ノード移動により2乗近似誤差が1/5程度に低減された．

強化学習への適用の評価に関しては，CMAC(Cerebellar Model Articulation Cotnroller) のような格子状に固定された関数近似方法との比較として，入力空間に格子状にノードを分布させ，固定したままの場合と，ノード移動や逐次的なノード追加を行った場合との比較を行った．また，ノードの追加を行うことは関連研究との対比では自律的な状態分割や基底の動的追加に対応しているが，「複雑度 (関数値) の大きなノードの近傍にノードを追加する」という追加指標が有効に機能することを示した．強化学習例題には水たまり問題と1自由度振子振

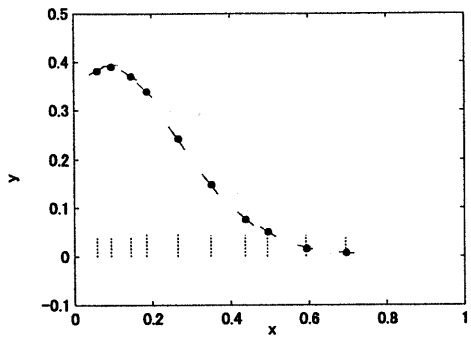

図 3: ノード移動後

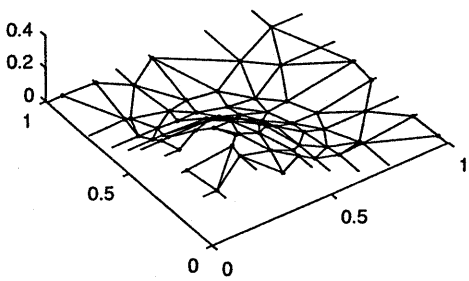

図 4: ノード移動後の分布

り上げ問題を取り上げ，各場合において，ノード数の変化，ノード移動の有無，ノードの逐次的追加による学習効率の変化を検証した．

表 1: 各学習例題での報酬積算値の比較

例題	ノード数	条件	報酬積算値
Puddle World	49	固定	-154.44
	45	ノード追加・移動	-84.45
Pendulumn	256	固定	6.565
	210	ノード追加・移動	7.320

その結果，表 1 のような結果を得た．いずれの例においても，ノードの逐次的追加と移動の組み合わせにより同程度あるいはそれ以下のノード数でより大きい報酬積算値を得ていることがわかる．これより，ノードの移動と複雑度に基づく逐次追加が有効に機能していることが確認された．その他に得られた結果を含めて考察を以下に示す．

- 固定ノード数の場合，ノード数が多い方が学習効率は改善されるが，一定個数よりも多くなりすぎると逆に学習効率は低下する．これは分解能が高すぎると逆に必要データ数の増加が強化学習性能に悪影響を与える可能性を示唆しているものと考えられる．

- ノード移動により，学習効率が改善できることを示した．この改善の割合は水たまり問題よりも問題設定を変更した 2 次元環境探索問題においてより顕著に表れた．このことから，ノード移動による適応的な分解能の変更が

強化学習において有効であること，および勾配変化の表現効率が報酬値に影響を与えやすい問題ほどノード移動の効果が高くなることが確認された．

- ノードの逐次追加によって，学習効率が改善できることを示した．逐次追加を行う場所の判別に関数値 $f(u)$(複雑度) を用いることで，効率の良いノードの追加が可能であることを示した．

以上により，提案する関数近似手法が強化学習にとって有効であることを確認した．