

要約

動径基底関数ネットワーク（Radial Basis Function Network: RBFN）は、任意の非線形関数に対し、基底関数を足し合わせることで関数近似を効果的に行うことができるニューラルネットワークのうちの一つであり、様々な問題に適用されている。しかし、RBFNには冗長なニューロンの存在や、基底関数が不足していることによって関数近似ができない場合があるといった問題点がある。このような問題に対して、ニューロン間の競合や、基底関数の複製を考慮したRBFNが提案されている。そこで本研究ではまず、提案されているRBFNの有効性を数値実験で確かめる。そして、強化学習の学習アルゴリズムの一つであるQ学習に適用することで、従来のRBFNと比較して有効であることを確認する。

キーワード：RBFN, シナプス可塑性方程式, 基底関数の複製, 強化学習, 学習の高速化

1 はじめに

関数近似問題やパターン識別に適したニューラルネットワークの一つにRBFNがある。RBFNは階層型ニューラルネットワークに比較してニューロンごとの局所的な学習が可能であるなどの優れた点を持つ。しかし、RBFNでは未知の非線形関数を近似するため、あらかじめ必要なニューロン数が不明であり冗長なニューロンを必要とする。一般に、ニューロンの増加は学習の遅延化や過学習の問題を生じることが知られている。

これらの問題を解決するために、適者生存型学習則に基づいたシナプス可塑性方程式を適用した、競合動径基底関数ネットワーク（Competitive RBFN: CRBFN）が提案されている[1]。CRBFNは、シナプス結合荷重間に競合を生じさせ、学習に必要なニューロンのみが自然に生き残るようになっており、冗長なニューロンの削減を図ることができる。しかし、CRBFNでは基底関数を追加する機能は備わっておらず、基底関数の数が足りない場合は関数近似自体が不可能となる。そこで、CRBFNに基底関数を複製して追加する機能を加えた、複製・競合動径基底関数ネットワーク(Reproductive CRBFN: RC-RBFN)が提案されている[2]。

そこで本研究では、RC-RBFNが効率よく基底関数を削除あるいは追加するニューラルネットワークであることを示したのち、強化学習の学習アルゴリズムの一つであるQ学習に適用することで、従来のRBFNと比較して有効であることを確認する。

— 2 複製・競合を考慮した動径基底関数ネットワーク —

2.1 競合動径基底関数ネットワーク

ニューラルネットワークは大きく分けて、素子であるニューロン、それらを結合するシナプス、そして動作規則により構成される。なかでも、記憶にもっとも関係した情報処理は、シナプスにおいて行われているとされる。記憶には種々のものが考えられるが、本論文では短期記憶と長期記憶に着目し、短期記憶はニューロンの発火頻度、長期記憶は細胞膜の特性の変化により生じるものとする。シナプスの可塑性を記述する方程式は、これらの要因を含んだものとなっていなければならない。また、実際の生体では、シナプス結合の性質が興奮性、抑制性であるのかは送り出すニューロンにより決まる。この原理をDale則と呼ぶ。さらに、微小な領域では成長や活動に必要な神経成長因子（Nerve Growth Factor: NGF）は競合によってシナプスに摂取される。これらの事実もシナプス可塑性のモデル化において重要な要因であると考えられる。そこで、発火頻度や膜の特性変化を生じる物質の時間変化と、生理学的拘束条件であるDale則や微小な領域での競合を考慮したシナプス結合荷重の大きさの時間変化は以下の方程式に従うものとする(図1参照)。

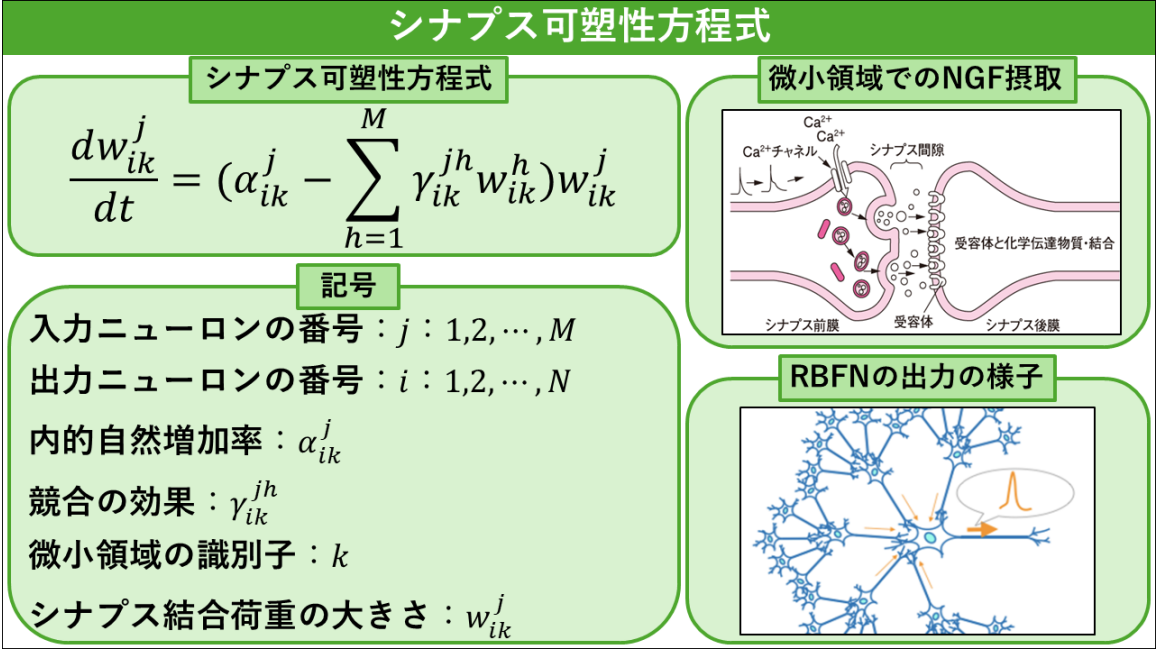


図1 シナプス可塑性方程式

このような、競合を考慮したシナプス可塑性方程式に従うRBFNがCRBFNである。また、方程式内のパラメータを調整することによって生存・分岐特性を示すことも示されている[4][5]。

2.2 ターミナルアトラクタ

先ほど述べたCRBFNでは、競合に負けたシナプス結合

荷重は平衡状態で0になる。しかし、平衡解への漸近は指数関数的に行われるので原理的には有限時間で平衡状態へ到達することはできない。そこで、望ましい出力が動径基底関数を定数倍して足し合わせることで実現できる特別の場合に、あるシステムが最終的に安定した学習結果に収束する、ターミナルアトラクタ(Terminal Attractor: TA)の概念を適用して、望ましい学習回数で平衡解へ収束するように修正されたシナプス可塑性方程式を以下のように定める(図2参照)。

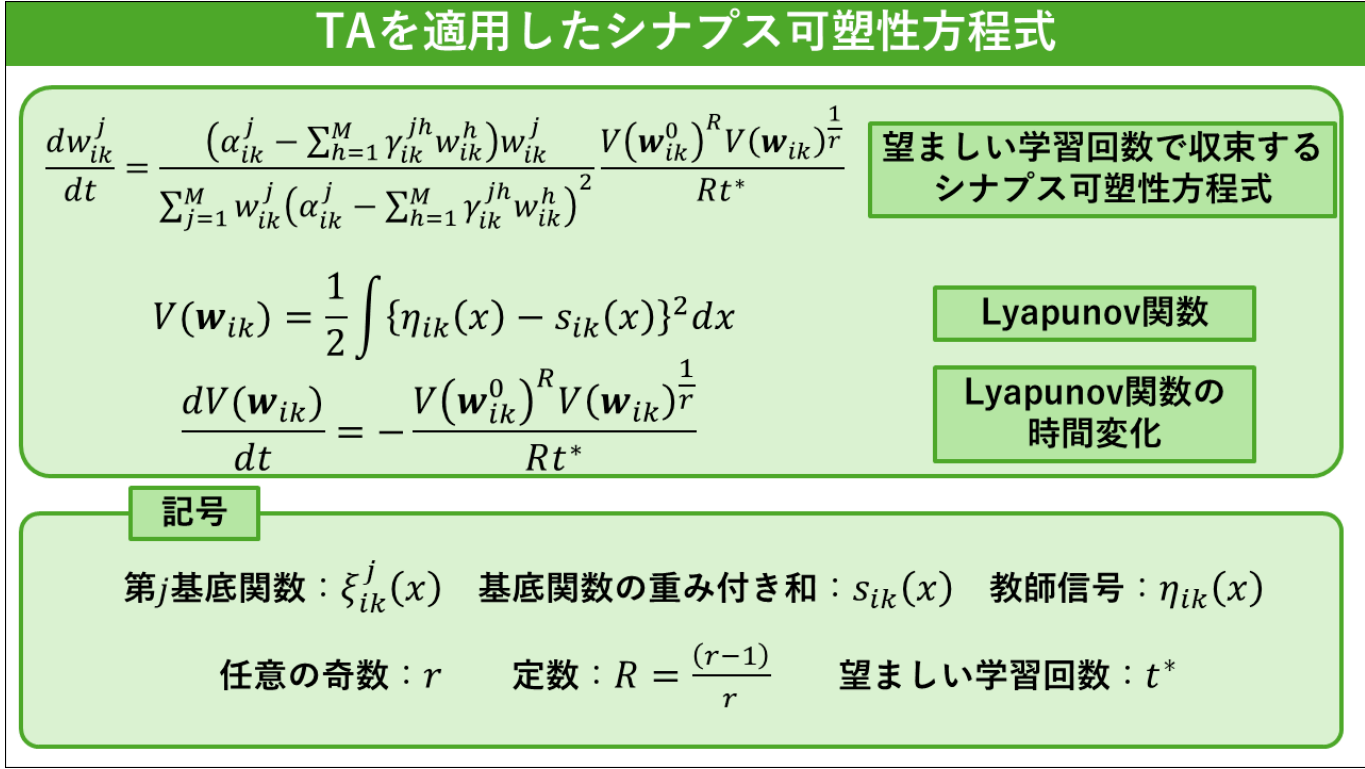


図2 TAを適用したシナプス可塑性方程式

ここで、Lyapunov関数は常に正であり、Lyapunov関数の時間変化は常に負である。したがって、このシステムは漸近安定であり、平衡点に漸近的に収束することが示される。

2.3 基底関数の複製

ここまでで、シナプス結合荷重間の競合を考慮したシナプス可塑性方程式を利用することにより、学習の効率化を図る手法が提案されてきた。ところが、CRBFNは冗長なニューロンを削除する能力をもつものの、必要とされる基底関数を追加する能力は備えていない。ニューラルネットワークに関数近似を行うために必要な数の基底関数が存在しない場合は、関数近似を行うこと自体が不可能となる。

そこで、新しい基底関数を追加する能力を備えたニューラルネットワークとしてRC-RBFNが提案されている。このRC-RBFNはシナプス可塑性方程式に関する考察から得られるものであり、必要な動径基底関数を効率的に追加することができる。新しい動径基底関数を複製するアルゴリズムを以下のように定義する(図3参照)。

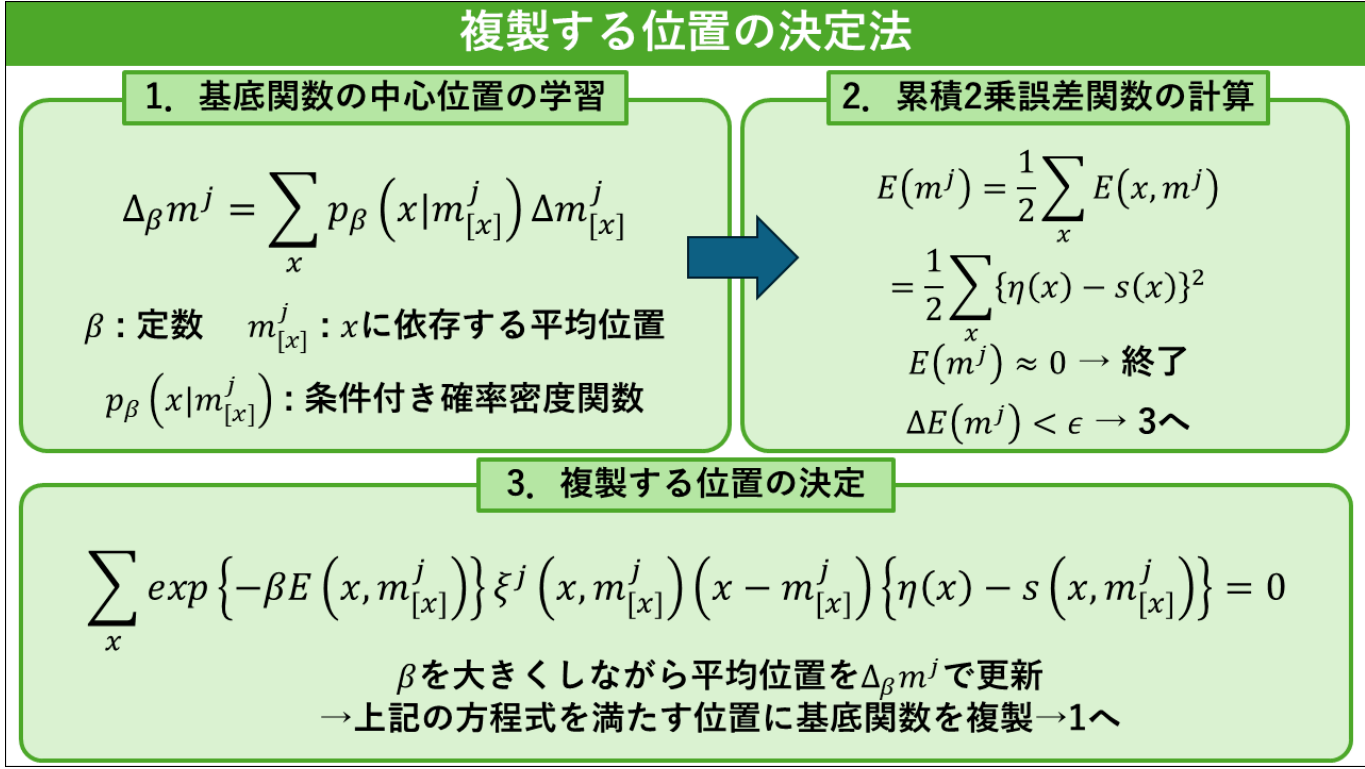


図3 複製する位置の決定法

また、この複製アルゴリズムにおける自由エネルギーは、データのクラスタリングのための手法であるメルティングと同様に定義されている[6]。メルティングとは、m = xかつβが∞である初期状態から、徐々にβを0へ近づけていきながら、パラメータmを自由エネルギーF_β(m)の最急降下方向に更新していくものである。その結果、パラメータmは徐々に同じ値をとりはじめ、最終的に一つの値に収束する。本研究の複製アルゴリズムではβを大きくしながらパラメータmを更新することでメルティングとは逆に、一つのmの値からx_i(v_i)に分岐させることで基底関数の複製を行っている。

— 3 強化学習とその手法について —

3.1 強化学習

強化学習は、エージェントが環境と相互作用し、行動の結果として得られる報酬をもとに最適な行動を学習する手法のことである。エージェントは状態に応じた行動を選択し、環境から報酬と次の状態を得て、そのフィードバックを用いて方策（ポリシー）を改善しながら報酬を最大化するように学習する。強化学習の手法は、主に価値ベース、方策ベース、およびアクター・クリティックの3種類に分かれ、連続的または大規模な状態空間に対しては、関数近似がよく用いられる。深層強化学習ではニューラルネットワークを使って関数近似を行い、複雑なタスクの学習を可能にしている。強化学習は、探索と利用のバランスを取りながら効率的に最適な行動方針を学ぶ点が特徴である。

3.2 Q学習

Q学習は強化学習におけるモデルフリー手法の一つで、状態行動価値（Q値）を用いて最適なポリシーを学習する。Q学習では、エージェントが環境と相互作用を繰り返しながら、各状態で最適な行動を選択する基準としてQ値を更新し、長期的な報酬を最大化する行動を学ぶ。この更新は、現在のQ値、得られた報酬、次の状態での最大Q値を考慮したQ値の更新式に基づいて行われ、エージェントが将来の報酬も含めた最適な行動を選択できるようになる。

3.3 RBFNを適用したQ学習

Q学習に対してRBFNは通常必須ではないが、組み合わせ使用されることがある。RBFNがもつ非線形関係を扱う能力は強化学習において、複雑な状態と行動の関係を効果的なモデル化を可能にするなどのメリットがある。しかし、局所的な基底関数を用いるRBFNでは基底関数がカバーしてない範囲の領域では関数近似を行うことが不可能という弱点も存在する。

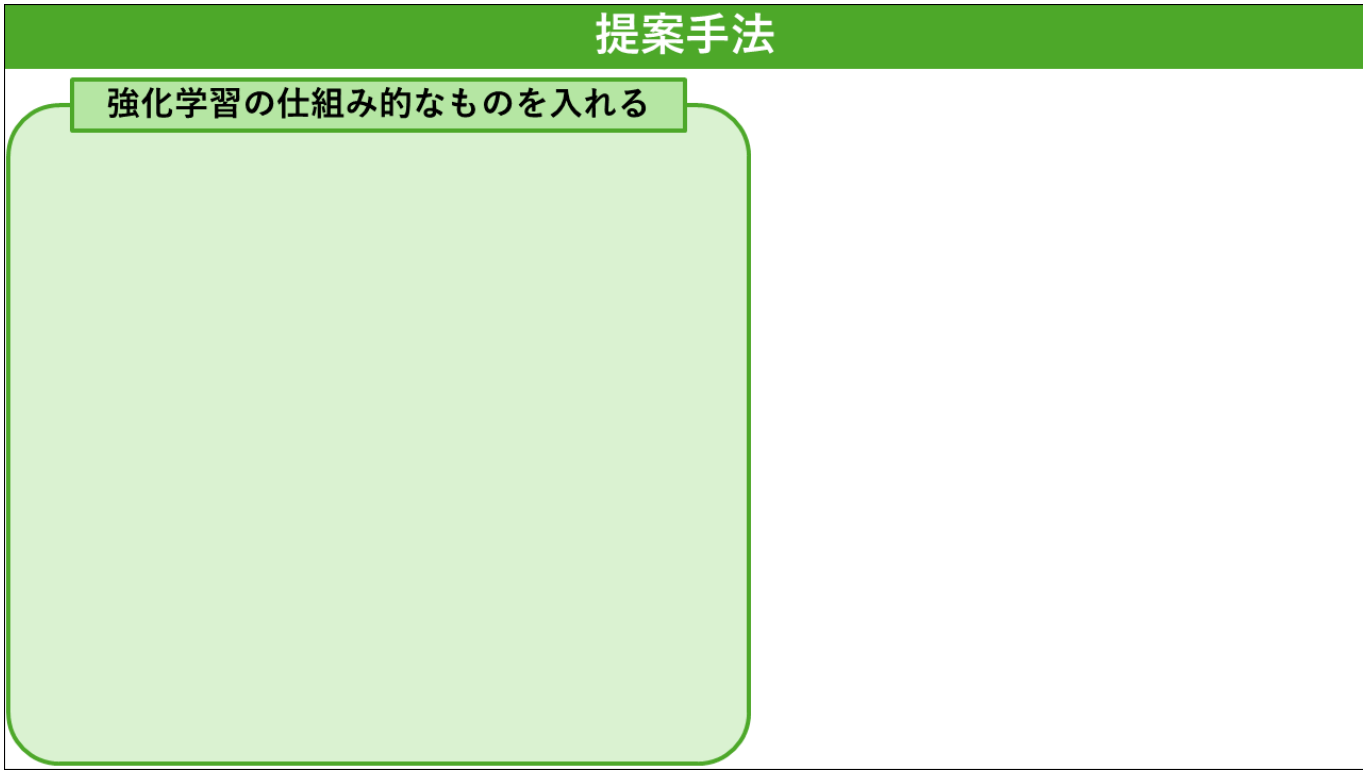


図4 なにか画像入れる

4 提案手法

本研究では、RBFNを発展させたニューラルネットワークとして提案されているRC-RBFNが実際に冗長なニューロンが削除していることや、基底関数の複製により教師信号との誤差を減少させることが可能であることを確認した。このRC-RBFNを強化学習の手法の1つであるQ学習に適用することで、RBFNを用いた場合と比較して有効であることを確認する。

適用する具体的な問題として、強化学習アルゴリズムの有効性を評価する際に一般的に使用される「車の山登り問題（Mountain Car Problem）」を選定する。この問題では、谷間に位置する車が、限られたエネルギーで山の頂上に到達することを目指して行動する。この課題は、エージェントが非線形な力学系に基づいて行動を選択する必要があるため、RC-RBFNを導入したQ学習アルゴリズムの効果を詳細に観察するための良い指標となると考えられる。

5 数値実験並びに考察

シナプス結合荷重の競合にターミナルアトラクタを適用した結果と基底関数の複製を行った結果を図4に示す。TA適用前のシナプス結合荷重の学習過程では、ほとんどのシナプス結合荷重が学習回数が80回以降は値の変化が見られず、収束したと判断した。それに対して、望ましい収束時間を50回で指定したTA適用後の学習過程では、適用前の収束した結果に学習回数が50回の時点で到達していることが分かる。TA適用前後のシナプス結合荷重の大きさを数値で比較してみると、w_10以外のシナプス結合荷重の差は0.1以下であり誤差であると考えられる。以上のことから学習回数を少なくしながら、同様の結果を得られていることが分かる。

基底関数の複製については、複製前は1個だった基底関数を図3の複製アルゴリズムによって複製することによって複製後は3個になっている。複製された基底関数の中心位置はそれぞれ教師信号の中心位置と一致しており、累積2乗誤差関数の遷移図からも誤差を減らせていることが確認できる。

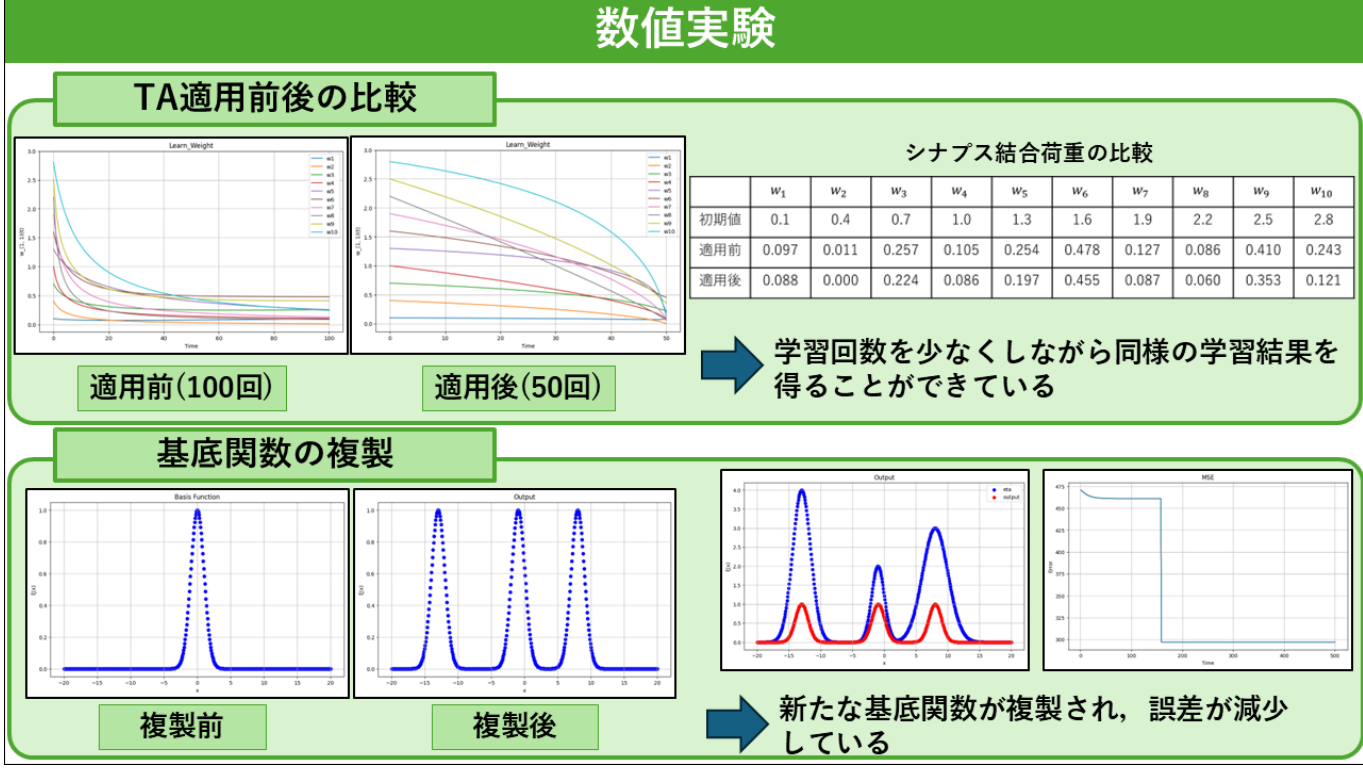


図5 TA適用前後のシナプス結合荷重と基底関数の複製

6 おわりに

本研究ではまず、RC-RBFNがRBFNでは行うことができなかった効率的な基底関数を複製と、冗長なニューロ

ンを削除が可能なことを確かめた．そして，強化学習の学習アルゴリズムの一つであるQ学習の関数近似の方法に，RC-RBFNを適用することを提案した．今後の課題としては，実際にQ学習に適用させて数値実験を行い有効性を確かめることや，車の山登り問題以外の問題でも実

験を行うことを考えている．

参考文献

[1] 奥原 浩之, 尾崎 俊治, “ 適者生存型学習則を適用した競合動径基底関数ネットワーク ”, 電子情報通信学会論文誌, pp. 3191-3199, 1997

[2] 奥原 浩之, 佐々木 浩二, 尾崎 俊治, “ 環境の変化に適応できる複製・競合動径基底関数ネットワーク ”, 電子情報通信学会論文誌, pp. 941-951, 1999

[3] 土屋 和雄, “ 複雑系の動力学とその機能 ”, システム制御情報学会誌 システム/制御/情報, Vol.39, No.1, pp.29～34, 1995

[4] Alexander S. Mikhailov, “ Foundations of Synergetics I: Distributed Active Systems (Springer Series in Synergetics) ”, 1994

[5] Wong, Y.-. fai . Clustering Data by Melting. Neural Computation 1993, 5 (1), 89–104. <https://doi.org/10.1162/neco.1993.5.1.89>.