社団法人 電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.
NC2000-166(2001-03)

連続値入力強化学習における Gauss-Sigmoid ニューラルネットワークの有効性

前原 伸一　杉坂 政典　柴田 克成

大分大学工学部
〒870-1192 大分市旦野原 700

E-mail: †{maehara,msugi,shibata}@cc.oita-u.ac.jp

あらまし 主として行動のプランニングに用いられてきた強化学習は、ニューラルネットワークと組み合わせることにより、センサからモータまでの一連の処理を総合的に学習することが可能となる。しかし、この組み合わせは Boyan らにより、学習を不安定に導くと指摘された [1]。これに対し、RBF ネットワークなどの局所的な情報表現の使用が有効であることが示されているが、これらの方法は、大域的な情報を表現する手段を持っておらず、その分、汎化能力が劣る。本稿では、RBF ネットワークの出力をシグモイドユニットの入力として用いる Gauss-Sigmoid ニューラルネットワーク（NN）を強化学習に用い、Boyan らが用いた hill-car 問題に適用した。その結果、シグモイド型 NN と比較して非線形関数近似能力が優れ、シグモイドユニット を用いても安定した学習が行えること、RBF ネットワークと比較して、Gauss-Sigmoid NN が学習を通して大域的な情報表現を獲得し、その上で汎化が有効に働く可能性を示した。

キーワード 強化学習、Gauss-Sigmoid ニューラルネット、RBF、Hill-car 問題 、汎化

Effectiveness of Gauss-Sigmoid Neural Nerwork in Reinforcement Learning with Continuous Inputs

Shin'ichi MAEHARA, Masanori SUGISAKA, and Katsunari SHIBATA

Depertment of Electrical and Electronic Engineering, Oita University
700 Dannoharu, Oita, 870-1192 Japan

E-mail: †{maehara,msugi,shibata}@cc.oita-u.ac.jp

Abstract Boyan et al. has pointed out that the combination of reinforcement learning and Sigmoid-based neural network sometimes leads to instability of the learning. In this paper, it is proposed that a Gauss-Sigmoid neural network, in which continuous input signals are put into a Sigmoid-based neural network through a RBF network,is utilized for reinforcement learning. It is confirmed through simulations of the hill-car task that the learning results are far better in the case of the Gauss-Sigmoid neural network, than in the case of the Sigmoid-based neural network. Also, it is comfirmed that the Gauss-Sigmoid neural network obtains a global representation in the hidden layer through learning and the generalization can work effectively on the representation.

Key words Reinforcement learning, Gauss-Sigmoid Neural Nerwork, RBF,Hill-car problem, generalization

1. まえがき

近年、自律ロボットや学習機械の開発などにおいて、強化学習の自律学習能力が注目を集めている。この自律学習能力は、強化学習が常に環境とのフィードバックループを利用しながら学習を行う枠組みから生み出される。従来、強化学習は、行動などのプランニングの学習としてとらえられており、予め設計された状態空間から各行動へのマッピングを学習することが一般的であった。しかし、強化学習はNNと組み合わせることにより、単に連続値入出力に対応できるだけでなく、センサからモータまでの、認識等も含めた一連の処理を総合的に学習することが可能となる[2]。また、NNの中間層が連続値状態空間の役割を果たし、これを他のタスクとの間で共有することで効率的に学習することも可能になると考えられる。

一方、Boyanらは、推力が小さな車が反動をつけて山を登るhill-car問題などを例として、NNと強化学習の組み合わせは学習の不安定につながることがあると指摘した[1]。これに対し、GordonやSuttonはCMAC、k-nearest-neighbor, RBF(Radial Basis Function)などの大域的な連続値入力信号を局所化する方法によって学習の不安定性を回避できることを示した[3][4][5]。タスクに強い非線形性が要求される場合、局所化された入力信号やテーブルルックアップのような表現は非常に有効である。しかし、このような関数では、出力は局所化された信号の線形和であらわされ、大域的な情報を表現する役割を持つ中間層を持っていない。さらに、近似精度を上げるためにRBFユニットの数を増やすと、情報表現がより局所的になるため、逆に汎化能力が低下してしまうというジレンマが存在する。このような中間層がないと、たとえば、ロボットに複数のタスクを学習させる際に、最初に学習させたことを次の学習に利用することはできず、空間認識のようにタスク間で共通に使えるものがあっても、1から学習し直さなければならない。また、正規化ガウス関数ネットワーク[6]はRBFの汎化能力を改善したものといえるが、やはりRBFユニットが密な領域では汎化能力は改善されない。

筆者らは、RBFとシグモイド型NNを組み合わせたGauss-Sigmoid NN[7][8]を強化学習に適用することを提案し、hill-car問題において、ランダムに選択した状態から1タイムステップのみの動作の学習を繰り返すことによって、1回の反動で斜面を登る動作を獲得させた。そして、シグモイド型ニューラルネットよりもよい動作を獲得できたことを確認した[9][10]。本稿では、何度も反動をつけないと登れないような問題を仮定するとともに、状態遷移を連続的に取り扱う。そして、シグモイド型NNに対する優位性とともに、RBFに対する優位性を検証する。

2. Hill-car問題

本論文では、タスクとして強い非線形関数近似を必要とするhill-car問題を用いる。このタスクは、車が斜面のある地点に置かれ、右の斜面の頂上に登った時のみ報酬がもらえるというものである。図1にhill-car問題を示す。

この図の斜面の式は

$$y = (x - 0.5)(x + 0.5), -0.5 < x < 0.5 \quad (1)$$

$$y = (|\,x\,| - 0.5)/\sqrt{1 + 5(|\,x\,| - 0.5)^2} \quad (2)$$
$$, |\,x\,| \geq 0.5$$

と表され、

$$\frac{dv}{dt} = (\frac{action}{m} - \frac{y'}{\sqrt{1 + y'^2}}g)/\sqrt{1 + y'^2}, \quad (3)$$

$$\frac{dx}{dt} = v, \frac{dy}{dx} = y', \quad (4)$$

という2つの微分方程式で車の運動を記述する。ただし、x, yは車の位置、vは車のx方向の速度、mは車の質量、$action$は車の推力、gは重力である。

この問題では、車が斜面を登ろうとする力が弱いため、傾斜が急な部分では、車の加速度は常に斜面を下る向きとなる。したがって、車が斜面のくぼみ$(x, v) = (0.0, 0.0)$から出発したとき、たとえ車が最大推力で一気に斜面を登ろうとしても頂上に到達することはできない。車が、一気に右の斜面を登ることができるかどうかの境界部分より右側に置かれた

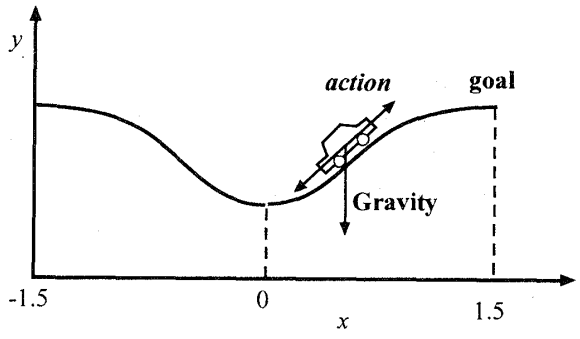

図1 Hill-car問題

ならば、車が出力すべき推力は右方向であり、状態評価値も大きい。しかし、車がその境界より少しでも左側に置かれたならば、車はいったん左の斜面を登った後に右側の斜面を登らなければならないため、出力すべき力は左方向となり、評価値も小さくなる。このように、この境界部分においては理想的な状態評価関数と出力すべき推力は共に不連続となり、その近似には強い非線形性が要求される。

3. Gauss-Sigmoid ニューラルネットワーク

まず、RBF ネットワークとその汎化能力を改善した正規化ガウス関数ネットワークについて説明する。RBF ネットワークの出力は次式のようにガウシアンで局所化した信号の線形和として表される。

$$output = \sum_{i=1}^{n} w_i g_i(x) + \theta \tag{5}$$

$$g_i(x) = exp(-\frac{1}{2}\sum_{d=1}^{D}\frac{(x_d - \mu_{i,d})^2}{\sigma_{i,d}^2}) \tag{6}$$

ただし、$(\mu_{i,1}, \cdots, \mu_{i,D})$ は i 番目の RBF ユニットの中心、$(\sigma_{i,1} \cdots, \sigma_{i,D})$ は i 番目の RBF ユニットのサイズ、θ はバイアス、n は RBF ユニットの個数、D は入力パターンの次元数である。RBF ネットワークの汎化能力はテーブルルックアップよりも優れていると言えるが、出力が局所的な情報の線形和で表さるという構造のため、シグモイド型 NN よりも大きく劣る。これに対し、RBF ネットワークの汎化能力を改善した正規化ガウス関数ネットワークでは、各 RBF ユニットの出力は、全ての RBF ユニットの出力の和で正規化され、

$$output = \sum_{i=1}^{n} w_i b_i(x) + \theta \tag{7}$$

$$b_i(x) = g_i(x) / \sum_{j=1}^{n} g_j(x) \tag{8}$$

と表される。この場合、各RBF ユニットが離れている場合には汎化が有効に働くが、密に配置された場所では汎化能力は改善されない。

一方、シグモイド型 NN では、出力関数がシグモイド関数で構成されるため、大域的な表現が可能である。しかし、シグモイド型 NN ではステップ関数などの強い非線形関数近似には適しておらず、Boyan らの指摘の通り学習が不安定になる場合がある。

そこで、本論文では図 2 に示すように RBF ユニットの出力をシグモイド型 NN の入力とする Gauss-Sigmoid NN [7][8] を用いる。シグモイド関数は強い非線形性の関数近似が苦手であるが、入力として局所化された信号を用いることによって、入出力間の非線形性が緩和され、強い非線形性を容易に実現することができる。さらに、学習を通して局所化された信号をシグモイド関数を出力関数とする中間層で統合することにより、大域的な表現を獲得することが可能となる。

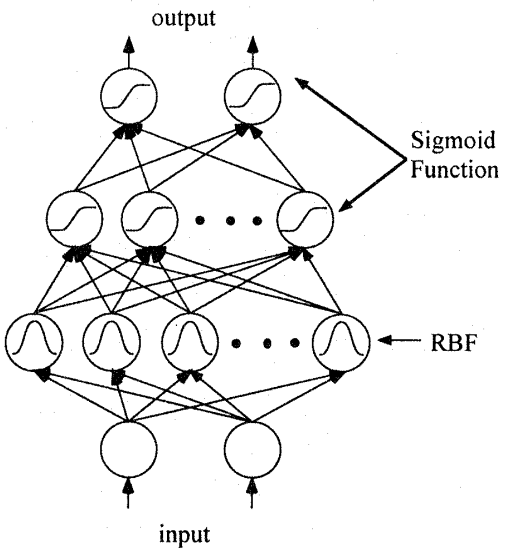

図 2　Gauss-Sigmoid NN

学習は、いずれも誤差逆伝搬法（BP 法）に基づいて行う。ただし、RBF ユニットに対して BP 法を直接適用すると、サイズ σ が小さくなり過ぎることがある。すると、σ や μ の更新量が大きくなり RBF ユニットの位置が大きく変化して学習が不安定になる。そこで、σ を log スケールで取り扱う。つまり、

$$\sigma_{j,d} = \exp(s_{j,d}) \tag{9}$$

とし、s のスケールで学習を行う。この時

$$\frac{1}{\sigma}\Delta\sigma = \Delta s \tag{10}$$

であるから、σ が 0 に近づくと s の空間は拡大され、誤差曲面の勾配は小さくなる。そこで、

$$\frac{d\sigma}{ds} = \exp(s) = \sigma \tag{11}$$

を、σ 、μ の更新量にそれぞれ用いる。RBF ユニット中心 μ の更新式を以下に示す。

$$\Delta\mu_{j,d} = -\frac{\partial E}{\partial \mu}\frac{\partial \sigma}{\partial s} = g_j(x)\frac{x_d - \mu_{j,d}}{\sigma_{j,d}}\delta_j \tag{12}$$

ここで、δ_j は伝搬誤差、E は出力の 2 乗誤差である。サイズ σ の訓練においては、まず s を

$$\Delta s_{j,d} = -\frac{\partial E}{\partial \sigma}\frac{\partial \sigma}{\partial s} = g_j(x)\frac{(x_d - \mu j,d)^2}{\sigma_{j,d}^2}\delta_j \quad (13)$$

にしたがって更新し、その後、s を (9) 式によってサイズ σ に変換して用いた。本稿では、RBF ネットワークも上記のように学習を行った。

4. Actor-critic アーキテクチャ

強化学習の主な手法として、Q-learning と Actor-critic アーキテクチャが知られているが、ここでは、Boyan らの論文と同様に、連続値動作を扱える Actor-critic アーキテクチャを用いる。Actor-critic アーキテクチャは Actor（動作生成部）と Critic （状態評価部）から構成され、Critic では過去の経験をもとに現在の状態の評価を行い、Actor ではより高い評価値の状態へ移動するための動作信号を学習する。Critic では、TD 誤差 $\hat{r}$

$$\hat{r} = r_t + \gamma P(x_t) - P(x_{t-1}) \quad (14)$$

を減少させるように、1 単位時間前の評価値を現在の評価値を用いて、次式を誤差として BP 法で学習していく。

$$\Delta P(x_{t-1}) = \alpha_p \hat{r}_t \quad (15)$$

ここで、γ は割引率、r は報酬、x は入力、$P(x_t)$ は評価値、α_p は学習係数である。一方、Actor では、出力 $a(t)$ に対し、$a(t)$ を中心とした確率分布から選ばれた $\tilde{a}(t)$ を実際の *action* 信号とした。その後、Actor では、状態評価値が大きくなるように、次式を誤差として BP 法で動作を学習をする。

$$\Delta a(t-1) = \alpha_a(\tilde{a}(t-1) - a(t-1))\hat{r}_t \quad (16)$$

ここで、α_a は学習係数、$\tilde{a}(t-1) - a(t-1)$ は試行錯誤量を表す。ここでは、Actor と Critic の 2 つを 1 つの Gauss-Sigmoid NN で構成する。そして、出力ユニットを 2 つ設け、一つを動作、一つを評価として取り扱う。動作出力がベクトルの場合は、その要素数分だけ動作用の出力を用意する。

5. シミュレーション

5.1 問題設定

ここでは、hill-car 問題を Actor-critic 型の強化学習によって学習する際に、1)Gauss-Sigmoid NN 、2)RBF network 、3) シグモイド型 NN で学習能力を比較する。ここで、車の初期値は山のくぼみ $(x, v) = (0.0, 0.0)$ とし、右の斜面の頂上に登ることを学習させる。Gauss-Sigmoid NN およびシグモイド型 NN では値域が -0.5 から 0.5 のシグモイドユニットを使用し、出力の -0.4 から 0.4 が、評価では 0.0 から 1.0、動作では -1.0 から 1.0 になるように線形変換を用いて出力した。また、学習前には中間層-出力層間の結合の重み値をすべて 0 とした。したがって、ネットワークから出力される評価値は常に 0.5 となり、動作も常に 0 となる。RBF ネットワークにおいても、RBF ユニットと出力層の結合の重み値は初期値 0 とし、出力層はバイアスを用いなかったため、学習前の出力は 0 となる。そこで、あらかじめ評価値の出力に 0.5 を加えることで学習を行った。

状態遷移は、(3)(4) 式をルンゲクッタ法を用いて計算する。Critic では、(15) 式にしたがって評価値を更新していくが、車が左側に飛び出した、もしくは丘の頂上に到達した場合は (14) 式の $P(x_t)$ を 0 とした。また、車の速度は 4.0 を越えた場合は 4.0、-4.0 より小さくなった場合は -4.0 に固定した。車の推力 *action* を $-1.0 \leqq action \leqq 1.0$ に固定しため、車が斜面のくぼみから頂上に到達するためには何度も左右交互に推力を出し、最終的に $v = 0$ で $x \leqq -0.69$ の状態まで登って反動をつけなければならない。Gauss-Sigmoid NN は中間層は 2 層で、1 層目 40 ニューロン、2 層目 10 ニューロンである。Gauss-Sigmoid NN と RBF ネットワークのガウシアンの個数は 400(20×20) とした。また、ここでは Gauss-Sigmoid NN と RBF ネットワークともに RBF ユニットの学習を行った。学習係数は Gauss-Sigmoid NN ではシグモイドユニット $100/\sqrt{(\text{結合しているユニット数})}$ 、RBF ユニット 0.5 、RBF ネットワークでは線形部 0.2、RBF ユニット 0.05 とし、慣性項は使用しなかった。学習回数は 2000 とした。

5.2 結　　果

Hill-car タスクにおける状態評価値と推力の分布、車の軌跡の様子を図 3 に示す。ただし、シグモイド型 NN では斜面のくぼみからゴールへ到達するという動作を獲得することができなかった。

車は斜面のくぼみから出発し、何度もくぼみ周辺をゴールにたどり着くことなく試行錯誤により行ったり来たりしていると、その部分の評価は下がり、相対的に車の通っていない状態の評価が高くなる。車はより評価の高い状態へと進もうとするため、行動の範囲は斜面の高い位置へと徐々に広がっていき、最終的には右の斜面の頂上であるゴール $x = 1.5$ もしくは

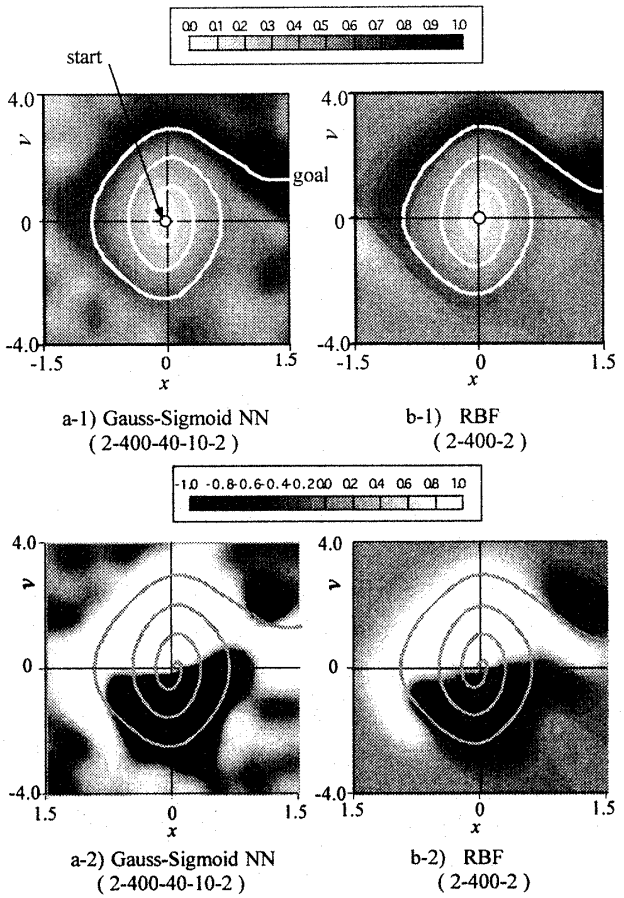

図3　状態 (x, v) に対する 1) 評価値、2) 推力の分布と車の軌跡

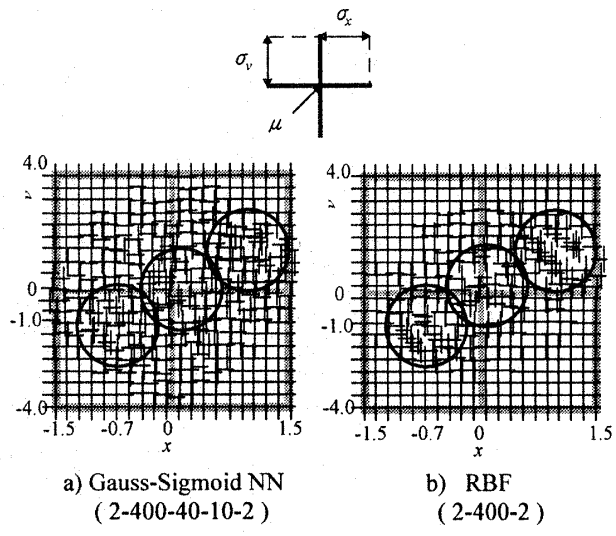

図4　学習後の各 RBF ユニットの配置

左の斜面の頂上にたどりつく。そして、車は $x \leq 1.5$ に達した場合のみ報酬 $r = 1.0$ が与えられ、その報酬が時間の経過とともにくぼみ周辺へ伝わり、試行錯誤のみであった動作が学習されていく。

図3の上の図から、Gauss-Sigmoid NN、RBF ネットワークともにゴールに一気に登れるか登れないかの境界部分では状態評価値が急激に変化していることがわかる。また、ゴールからスタート地点まで、反時計回りに評価値の尾根が渦をまいているのが確認でき、車の軌跡はその尾根に沿っている。学習が進行する様子を観察すると、学習前の出力層への結合の重み値は両者とも 0 であり、報酬の影響も少ないため、学習初期において得られる推力は小さい。したがって、車が一往復の間で登れる高さは非常に小さい。また、明らかに登れない状態でも登る方向への推力を出す傾向にあるため、軌跡はひし形に近い形になる。その後、学習が進むにつれて評価値の尾根は次第にスタート地点へと渦をまいていき、動作の学習が進むためゴールへ到達までの行動回数は急激に少なくなる。最終的な車の軌跡は、まず右向きに進んだ後、左右に動きながらスタート地点を通過する速度を増していき、4回目に右方向に進んだ後ゴールに到達することがわかる。RBF ネットワークでは、車が通っていない部分での評価値はほぼ初期値 0.5 のままであるが、Gauss-Sigmoid NN では様々な値が出力されている。Gauss-Sigmiod NN は入力層と中間層の間に小さな乱数で初期値が決定される重み値を持つため、その影響が出力層に出たためと考えられる。

推力に注目すると、Gauss-Sigmoid NN、RBF ネットワークともに推力の方向の変化は明確である。これは、小さい推力は効率的でないため、多くの状態で左右にほぼ最大推力で車が進んでいることを表している。また、推力の分布の様子も Gauss-Sigmoid NN と RBF ネットワークの間でよく似ている。左右の推力の境界は、多少右上がりになっているが、これは左右方向の速度が 0 なる前に反対向きの推力が働いていることを示している。

次に、学習後の RBF ユニットの配置を図 4 に示す。RBF ユニットは、図中の丸で示した、速度 0 の状態から右にかけ登るスタート地点周辺、左側に飛び出さないようにかつ、右の頂上へスムーズに登るために左方向に対する速度にブレーキの役割をする $(x, v) = (-0.7, -1.0)$ 周辺、そして、もう一度左へ戻るかそのまま一気に頂上に登るかというゴール周辺、つまり、強い非線形性が要求されるところに集まる傾向がある。特に RBF ネットワークにその傾向が強い。

学習後に斜面のくぼみからゴールに到達するまでの車の x 方向の位置 x、速度 v、推力 $action$、評価

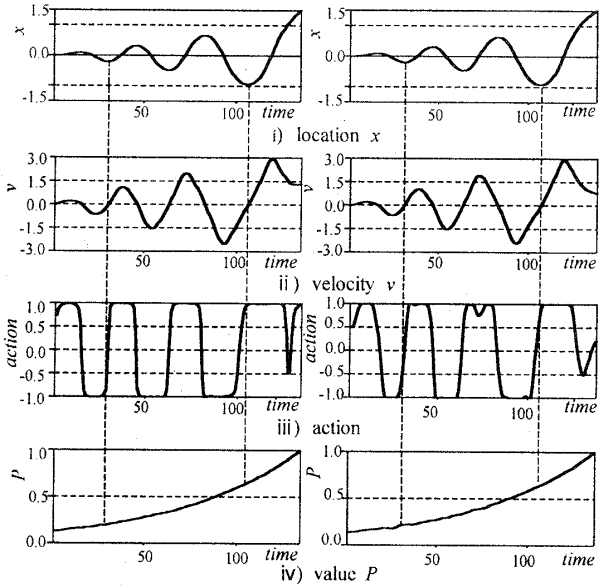

図5　学習後の各変数の時間変化

値 P の時間変化の様子を図5に示す。最初は右向きに力を出して少し山を登り、左、右と往復しながら斜面を登っていき、4回目に右側に登ったときに、十分な反動を得てゴールに到達している。推力は、はじめは速度が0になる時に0となり方向が変わっているが、徐々に速度が0になる前に方向を変え、4回目に右に進む前には、速度0で右向きに最大推力が出力されている。また、2つのネットワークとも、評価値はゴールに近づくにしたがってなめらかに上昇している。これは、車が評価値の尾根に沿って斜面を登っていることを示している。ゴールに到達する直前では車の出力する力は両者とも負になっているが、これは過去の学習においてそのまま一気に登れず、もう一度反対方向に進み、さらに反動をつけて登ろうと学習した結果であり、ゴール直前のため、ステップ数には大きな影響がなかったために残ったものと考えられる。

図6に hill-car 問題 の学習曲線を示す。これはスタート地点からゴールに到達するまで要するステップ数が学習が進むにつれて減少する様子を表している。Gauss-Sigmoid NN と RBF には大きな違いは現れていない。学習初期には、車は斜面の周辺を小さい推力で行ったり来たりするため、多くのステップを必要とするが、何度がゴールにたどり着くと急速にステップ数は減少していく。

5.3　シグモイド型 NN との比較

前述のように、シグモイド型 NN では、山を登

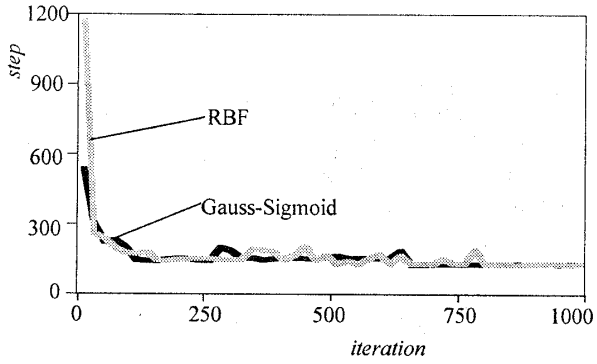

図6　学習曲線

ることができなかったので、より簡単な問題での結果を示す [9][10]。ここでも、RBF ネットワークと Gauss-Sigmoid NN の学習結果は、ほぼ同じであったため、Gauss-Sigmoid NN とシグモイド型 NN の結果を示す。車の推力を $-3.0 \leqq action \leqq 3.0$ とし、$-0.5 \leqq x \leqq 1.5$ 、$-4.0 \leqq v \leqq 4.0$ の範囲でランダムに選択した状態から、1タイムステップ後の動作の学習を繰り返した。Gauss-Sigmoid NN とシグモイド型 NN の中間層は、ともに1層で中間層の数は40、Gauss-Sigmoid NN の RBF ユニットは $110(11 \times 10)$ とした。状態評価値、推力の分布と車の軌跡を図7に示す。Gauss-Sigmoid NN では、スタート地点の左に、評価値の尾根が明確に表されている。また、左-右-左と一度左方向に行くだけでゴールに到達することができており、その軌跡は評価値の尾根に沿っている。これに対し、シグモイド型 NN では、その部分に明確な評価値の尾根は確認できない。また、最大推力が3.0であるにもかかわらず、何度も行ったり来たりしなければゴールに到達できない。また、Gauss-Sigmiod NN は、推力の変化する境界を明確に表すことができているが、シグモイド型 NN ではそのような境界は観察することができない。これらのことから、シグモイド型 NN の学習能力が劣るのは、非線形関数近似能力が不十分であることが原因だと考えられる。

以上のように、Gauss-Sigmoid NN は連続値入力強化学習において、シグモイドユニットを用いているにもかかわらず、シグモイド型 NN が近似できないような非線形の強い問題も近似することができ、その能力は RBF ネットワークとほぼ同等だといえる。これは RBF ユニットで情報を局所化しそれをシグモイド型 NN の入力として用いているためである。

5.4　大域的情報表現

次に、5.2節での初期値をくぼみ $(x, v) = (0.0, 0.0)$

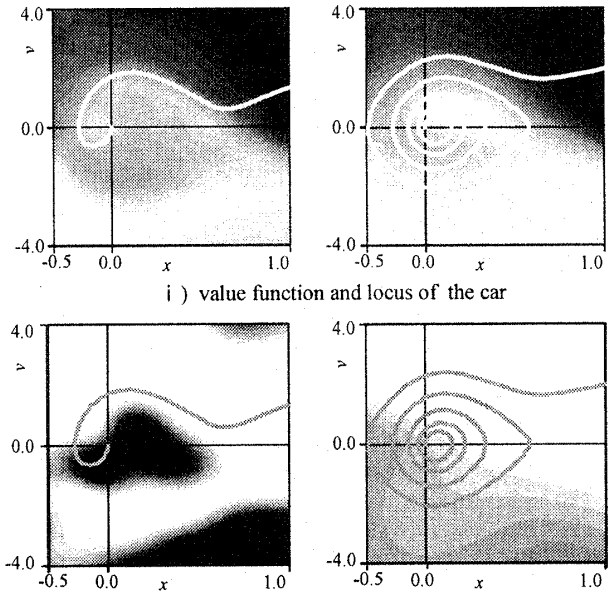

図 7　Gauss-Sigmoid NN とシグモイド型 NN の比較

とした 2000 回の学習で、Gauss-Sigmoid NN のシグモイドユニットがどのような内部表現を獲得しているか観察する。出力層に、中間層との結合係数 0 のニューロンを新たに追加し、そのニューロンに対して教師あり学習を行った。教師信号は $(x, v) = (0.0, 0.0)$ で 0.0 （図の小さい○）、 $x = 1.5$ 、 $v > 0.0$ （図の小さい長方形）で 1.0 とし、この中からランダムに入力を選んで学習した。学習後の x、v に対する出力の分布を図 8 に示す。図から、Gauss-Sigmoid NN では、中間層内部に以前学習した状態評価値の尾根の情報を蓄えていることがわかる。一方、RBF ネットワークでは、中間層を持たないため、教師信号があたえられた部分では学習されるものの、その領域は非常に小さく、ほとんどの領域で 0.5 となった。このように、Gauss-Sigmoid NN では、中間層内部に、以前学習した大域的な情報としての知識を蓄えることができる。したがって、このような知識が活用できる別のタスクを学習する場合、学習を 1 からやり直す必要はなく効率的な学習が行えることが期待される。

そこで、以前学習した大域的な情報としての知識を活用するタスクとして、くぼみを初期値とした学習後に、出力層との結合の重み値を 0 にして、もう一度同じゴールへ登らせるというシミュレーションを行った。しかし、この場合、逆に Gauss-Sigmoid

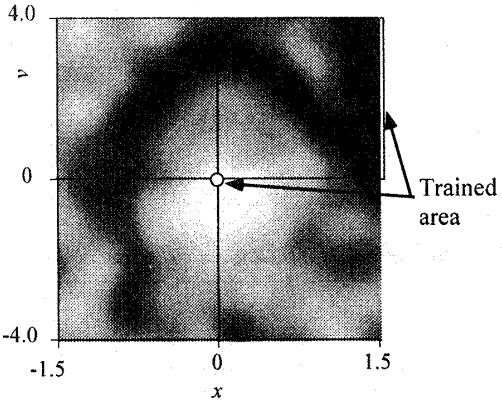

図 8　教師あり学習後の出力分布

NN では、車はスタート地点付近で止まってしまい、学習が進まなかったが、RBF ネットワークでは登ることができた。中間層に蓄えられた知識は、ゴールに一度たどり着き、報酬を得た後に初めて再現される。そのため、スタート地点周辺でゴールにたどり着くことなく行ったり来たりしているうちに、本来尾根になるべきところの評価が、逆に低い値になってしまったと考えられる。

そこで、より報酬を容易に獲得できるようにするため、2000 回の学習後、中間層-出力層間の結合の重み値をすべて 0 にし、$-1.5 \leqq x \leqq 1.5$、$-4.0 \leqq v \leqq 4.0$ の範囲でランダムに初期値を選び、そこからゴールにたどり着くことを学習させた。スタート地点からゴールまでの所要ステップ回数を図 9 に示す。Gauss-Sigmiod NN では図 8 に示したように、結合の重み値をリセットする前に学習した情報が蓄えられているため、評価値の尾根がスタート地点に伸びていくのが早く、一度ゴールに到達するとすると急速に所要ステップ回数が減少する。一方、RBF では、たとえランダムに与えられる初期値がゴールに近くてもスタート地点からは何度も行ったり来たりしなければ登れず、また評価値の尾根もなかなか伸びない。乱数の初期値をかえて 3 回シミュレーションを行ったが、いずれも場合も RBF ネットワークよりも明らかに Gauss-Sigmoid NN の方が少ないステップ数でゴールに到達できた。

図 10 に、学習によって得られた状態評価値と車の軌跡の様子を示す。Gauss-Sigmoid NN が結合の重み値をリセットする前と非常によく似た軌跡をたどっているのに対し、RBF ネットワークではスタート地点付近で何度も行ったり来たりしている。これは、Gauss-Sigmoid NN では、以前の学習で蓄えられた情報が、ゴールたどり着き報酬を得たことで素速

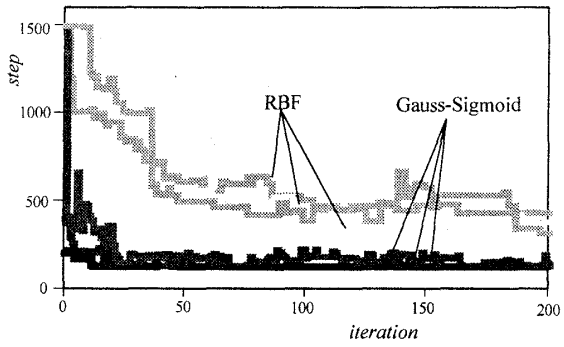

図 9　出力層への結合の重み値を 0 にリセットした後に再学習させたときの学習曲線

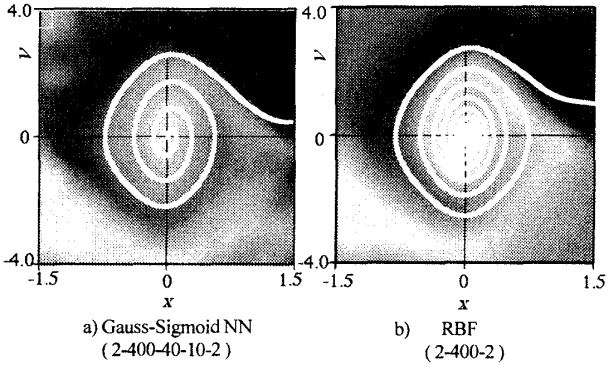

図 10　200 回の再学習を行ったときの状態評価値と車の軌跡の様子

く再現されたためと考えられる。また、RBF では、200 回の学習回数では不十分であったと考えられる。

6. 結　　論

本稿では、入力信号が大域的な連続値信号である場合に Gauss-Sigmoid ニューラルネットワークを用いることを提案した。Gauss-Sigmoid NN を非線形性の強い hill-car タスクにおいて適用した結果、Gauss-Sigmoid NN ではシグモイドユニットを有していても優れた非線形関数近似能力を示し、RBF ユニットと同等、シグモイド型 NN よりもよい学習能力を実現できるた。さらに、Gauss-Sigmoid NN は、一度学習した大域的な知識を中間層に蓄えることができるため、RBF ネットワークと比較して、優れた汎化能力を備えている一例を示した。しかし、タスクによってはその能力を十分引き出せない場合があるなど今後さらに研究を進めていく必要があると考えられる。また、どのような場合に Gauss-Sigmiod NN が有効であるかを示すことは今後の課題である。

文　　献

[1] Boyan, J.A. & Moore, A.W., "Generalization in Reinforcement Learning: safely approximating the value function", in Advances in Neural Information Processing Systems, Vol. 7, pp. 369-376, The MIT Press(1995)

[2] 柴田克成, 岡部洋一, 伊藤宏司, "Direct-Vision-Based 強化学習 -センサからモータまで", 計測自動制御学会論文集, Vol.37, No.2 (2001) in press

[3] Gordon, G. J. "Stable Function Approximation in Dynamic Programing", Proc. of the 12-th ICML, pp.261-268 (1996)

[4] Sutton, R. S. "Generalization in Reinforcement Learning : Successful Examinoles Using Space Coarse Coding", In Advanced in Neural Information Processing System, Vol8, pp.1038-1044 (1996)

[5] Sutton, R. S. & Barto, A. G. "Reinfoecement Learning", The MIT Press (1998)

[6] 森本淳, 銅谷賢治, "強化学習を用いた高次元連続状態空間における系列運動学習-起き上がり運動の獲得-", 電子情報通信学会論文誌, J82-D-II, No. 11, pp.2118-2131 (1999)

[7] Shibata, K. & Ito, K. "Gauss-Sigmoid Neural Network", Proc. of IJCNN'99, #747(1999)

[8] 柴田克成, 前原伸一, 杉坂政典, 伊藤宏司 "Gauss-Sigmoid ニューラルネットワーク", 第 13 回自律分散システムシンポジウム資料,pp.133-138 (2001)

[9] 前原伸一, 杉坂政典, 柴田克成, "Gauss-Sigmoid ニューラルネットワークを用いた強化学習の安定性", 計測自動制御学会九州支部第 19 回学術講演会予稿集, pp.475-478(2000)

[10] Maehara S., Sugisaka M., Shibata K. "Reinforcement Learning Using a Gauss-Sigmoid Neural Network", Proc. AROB 6th'01 vol2, pp.562-565 (2001)