

区間値遺伝子を用いた区間 ES によるニューロエボリューション Evolving Neural Networks by means of Interval Evolution Strategy with Interval-valued Genotypes

山下 彬 松瀬 高志 和田 哲也 岡田 英彦

Akira Yamashita, Takashi Matsuse, Tetsuya Wada and Hidehiko Okada

1. はじめに

これまでに、ユニット間結合強度やユニットしきい値が実数ではなく区間に拡張された階層型ニューラルネットが提案されている[1]. さらに、学習用データを用いて、この区間ニューラルネット (INN) の結合強度としきい値を学習させるために、従来のバックプロパゲーションアルゴリズムを拡張した手法も提案されている[1]. しかし、区間ニューラルネットの教師なし学習法は提案されていなかった. これに対して、近年、進化計算手法を用いたニューラルネットの教師なし学習法が研究されており[2], ニューロエボリューションと呼ばれている.

本研究では、解の遺伝子表現 (genotype) の値として区間値を用いることが可能な進化計算手法を提案している[3]. 本提案手法の有効性評価の一環として、提案手法の一種である区間 ES を、前記 INN のニューロエボリューションに適用した.

2. 区間ニューラルネット

本研究で用いるニューラルネットは、文献[1]で提案された INN である. この INN は、3 階層のフィードフォワード型であり、結合強度としきい値が区間値をとる. INN の構造を図 1 に示す. なお、簡単化のため、INN の出力層ユニット数は 1 とする.

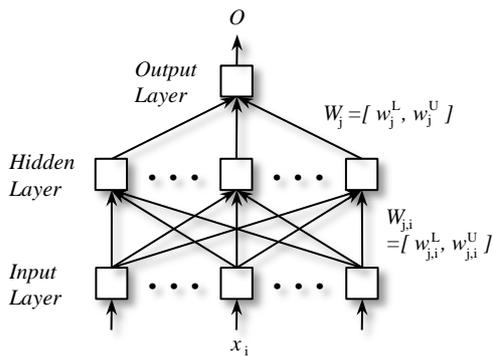


図 1: 区間ニューラルネットの構造[1]

この INN は、実数入力ベクトル \mathbf{x} を受け取り、区間出力値 O を、式(1)-(5)に基づいて計算する.

入力層 :

$$o_i = x_i. \quad (1)$$

中間層 :

$$Net_j = \sum_i W_{j,i} o_i + \theta_j, \quad (2)$$

$$O_j = f(Net_j). \quad (3)$$

出力層 :

$$Net = \sum_j W_j O_j + \theta, \quad (4)$$

$$O = f(Net). \quad (5)$$

式(1)-(5)において、 x_i, o_i は実数であり、 $Net_j, Net, W_{j,i}, W_j, \theta_j, \theta, O_j, O$ は区間である. ユニット内部関数 $f(x)$ にはシグモイド関数 $f(x) = 1/(1+e^{-x})$ を用いる. この $f(x)$ による区間入力値から区間出力値へのマッピングを図 2 に示す.

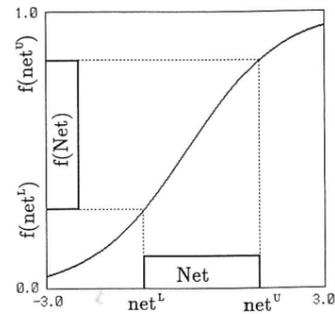


図 2: 中間層・出力層ユニットにおける入出力関係[1]

式(2)-(5)における区間演算は次の式(6)-(8)に基づいて行われる[4]. 2つの閉区間を、 $A = [a^L, a^U], B = [b^L, b^U]$ とする.

$$\begin{aligned} A + B &= [a^L, a^U] + [b^L, b^U] \\ &= [a^L + b^L, a^U + b^U]. \end{aligned} \quad (6)$$

$$\begin{aligned} k \cdot A &= k[a^L, a^U] \\ &= [ka^L, ka^U], \text{ if } k \geq 0, \text{ or } [ka^U, ka^L], \text{ if } k < 0. \end{aligned} \quad (7)$$

$$\begin{aligned} A \cdot B &= [a^L, a^U] \cdot [b^L, b^U] \\ &= [\min(a^L b^L, a^L b^U, a^U b^L, a^U b^U), \\ &\quad \max(a^L b^L, a^L b^U, a^U b^L, a^U b^U)]. \end{aligned} \quad (8)$$

図 1 の INN において、入力層ユニット数を n ($i=1,2,\dots,n$), 中間層ユニット数を m ($j=1,2,\dots,m$) とすると、 $W_{j,i}$ の数は mn , W_j の数は m であり、結合強度の数は $mn+m$ である. さらに、ユニットしきい値の数は $m+1$ である. したがって、INN に含まれる結合強度およびしきい値の総数は $mn+2m+1$ である. 本研究が提案する区間 ES は、この $mn+2m+1$ 個の区間値を genotype の値として用いる. genotype を $\mathbf{V} = (V_1, V_2, \dots, V_D)$ と表すと、 $D = mn+2m+1$ であり、 V_1, V_2, \dots, V_D は、対応する結合強度もしくはしきい値がとる区間値である. V_1, V_2, \dots, V_D は、それぞれ、上限と

下限, もしくは中心と幅によって規定できる. $V_i = [v_i^L, v_i^U]$ もしくは $V_i = (v_i^C, v_i^W)$ であり, $v_i^L, v_i^U, v_i^C, v_i^W$ は V_i の上限, 下限, 中心, 幅を表す.

3. 区間 ES

提案する区間 ES のプロセス構成は, 従来の ES と共通である (図 3). ただし, 解の初期化, 適合度評価, 子個体生成の方法が, genotype の区間値を扱うことができるように拡張される.

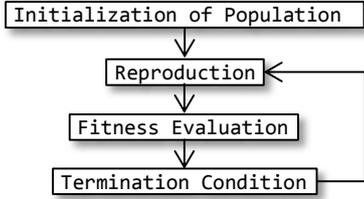


図 3: 区間 ES のプロセス構成

3.1 個体の初期化

個体の集団サイズを P とする. 初期化プロセスでは, 各個体の genotype V_1, V_2, \dots, V_P をランダムに初期化する. $V_p = (V_{p,1}, V_{p,2}, \dots, V_{p,D})$ であり, $V_{p,d}$ は区間である ($p=1,2,\dots,P, d=1,2,\dots,D$). すなわち, $V_{p,d} = [V_{p,d}^L, V_{p,d}^U]$ もしくは $V_{p,d} = (V_{p,d}^C, V_{p,d}^W)$ と表される. 前者を LU モデル, 後者を CW モデルと呼ぶことにする. V_p はニューラルネットの結合強度もしくはしきい値のため, $V_{p,d}$ の初期値としては, 絶対値があまり大きくない値が望ましいと考えられる. そこで, $V_{p,d}$ の初期値を, 平均 0 の正規乱数, もしくは中心 0 の一様乱数を用いて決定する. $V_{p,d} = [V_{p,d}^L, V_{p,d}^U]$ の場合, 乱数を 2 つ生成し, その小さい (大きい) ほうを $V_{p,d}^L$ ($V_{p,d}^U$) に代入する. 一方, $V_{p,d} = (V_{p,d}^C, V_{p,d}^W)$ の場合, 乱数を 2 つ生成し, その一方を $V_{p,d}^C$ に, もう一方の絶対値を $V_{p,d}^W$ にそれぞれ代入する.

3.2 子個体の生成

ある親個体を V_a とし, この親個体から生成される子個体を V_z とする. 集団から V_a を選ぶ方法は従来の ES と同一の方法を用いる.

LU モデルの場合は,

- $V_a = (V_{a,1}, V_{a,2}, \dots, V_{a,D}), V_{a,i} = [v_{a,i}^L, v_{a,i}^U]$
- $V_z = (V_{z,1}, V_{z,2}, \dots, V_{z,D}), V_{z,i} = [v_{z,i}^L, v_{z,i}^U]$

である. 子個体の $v_{z,i}^L$ と $v_{z,i}^U$ は次式によって決定する.

$$v_{z,i}^L = v_{a,i}^L + rand, \quad (9)$$

$$v_{z,i}^U = v_{a,i}^U + rand. \quad (10)$$

ただし, $rand$ は乱数を表しており, 平均 0 の正規乱数や中心 0 の一様乱数を用いる. また, $v_{z,i}^L > v_{z,i}^U$ となった場合は,

$v_{z,i}^L \leq v_{z,i}^U$ が満たされるように修正する必要がある. この修正方法として, 例えば次の 4 通りがあげられる: (1) $v_{z,i}^L \leftarrow v_{z,i}^U$, (2) $v_{z,i}^U \leftarrow v_{z,i}^L$, (3) $tmp \leftarrow v_{z,i}^L, v_{z,i}^L \leftarrow v_{z,i}^U, v_{z,i}^U \leftarrow tmp$, (4) $v_{z,i}^L \leftarrow (v_{z,i}^L + v_{z,i}^U)/2, v_{z,i}^U \leftarrow v_{z,i}^L$.

CW モデルの場合は,

- $V_a = (V_{a,1}, V_{a,2}, \dots, V_{a,D}), V_{a,i} = (v_{a,i}^C, v_{a,i}^W)$

- $V_z = (V_{z,1}, V_{z,2}, \dots, V_{z,D}), V_{z,i} = (v_{z,i}^C, v_{z,i}^W)$

である. 子個体の $v_{z,i}^C$ と $v_{z,i}^W$ は, LU モデルの場合と同様に, 次式によって決定する.

$$v_{z,i}^C = v_{a,i}^C + rand, \quad (11)$$

$$v_{z,i}^W = v_{a,i}^W + rand. \quad (12)$$

$v_{z,i}^W < 0$ となった場合は, $v_{z,i}^W \geq 0$ が満たされるように修正する必要がある. この修正方法として, 例えば $v_{z,i}^W \leftarrow 0$ とすることが挙げられる.

3.3 個体の適合度評価

genotype V_p に対応する INN の個体の適合度を評価するために, その INN を対象問題に適用して性能を測定する. 後述する実験では, 対象問題は区間関数 $y=F(x)$ の近似である. ただし, $F(x)$ は INN および区間 ES のプロセスに対して与えられず, さらに, 入力 x に対する INN の目標出力値も陽には与えられない. あるサンプリングされた複数の入力値に対して, INN の出力値と関数 $F(x)$ の値の誤差の総和だけが区間 ES に対して与えられる. 区間 ES の適合度評価プロセスにおいては, この誤差の総和を個体の適合度評価値として用い, 集団内の個体をランキングする.

4. 評価実験

INN の対象問題として区間関数の近似問題を取り上げ, 本研究で提案した区間 ES による INN の教師なし学習を実験的に試みる. 簡単化のため, 区間関数 $F(x)$ の入力 x は $0 \leq x \leq 1$ の実数とする. これにより, INN の入力層ユニット数は 1 である.

集団内の個体をランクづけするために, 各個体の適合度を次のように決定した. ある個体に対応する INN $\rightarrow, 0 \leq x \leq 1$ の範囲で一様ランダムにサンプリングされた入力値を与え (x_r とする), その入力値に対する INN の出力区間値を得る ($O_r = [O_r^L, O_r^U]$ とする). 一方, その入力値に対する目標関数 F の値も求める ($F(x_r) = [F_r^L, F_r^U]$ とする). この O_r と $F(x_r)$ の誤差 e_r を, $e_r = (y_r^L - f_r^L)^2 + (y_r^U - f_r^U)^2$ と定義する. x_r をランダムに 100 回サンプリングして誤差 e_r の総和を求め, 個体の適合度として用いる. 前記の通り, 区間 ES に対してフィードバックされる値はこの誤差 e_r の総和のみであり, $F(x_r), e_r$ のいずれの値も区間 ES には与えられない.

本実験では例題として 3 種類の目標関数を用いた. これらの関数は次の通りである. なお, 目標区間関数 $F(x)$ の下限および上限をそれぞれ $F(x)^L$ および $F(x)^U$ とする.

例題(1)

$$F(x)^L=0.2\sin(2x)+0.1x^2+0.3, F(x)^U=0.2\sin(2x)+0.1x^2+0.6$$

例題(2)

$$F(x)^L=0.1\sin(2x)-0.1x+0.4, F(x)^U=0.1\sin(2x)+0.1x+0.6$$

例題(3)

$$F(x)^L=0.1\sin(2x)+0.1x+0.3, F(x)^U=0.1\sin(2x)-0.1x+0.6$$

これらの区間関数を図 4-6 に示す.

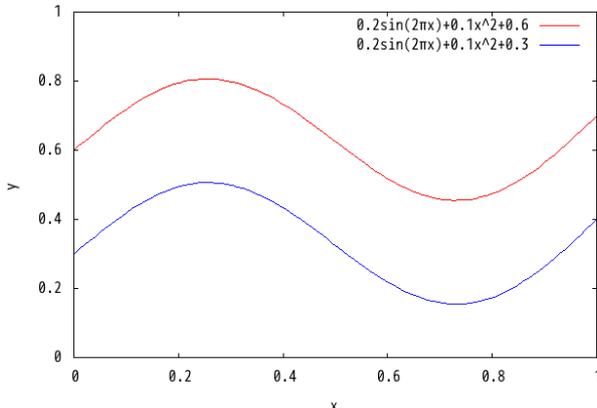


図 4: 本実験において INN が近似する目標区間関数(1)

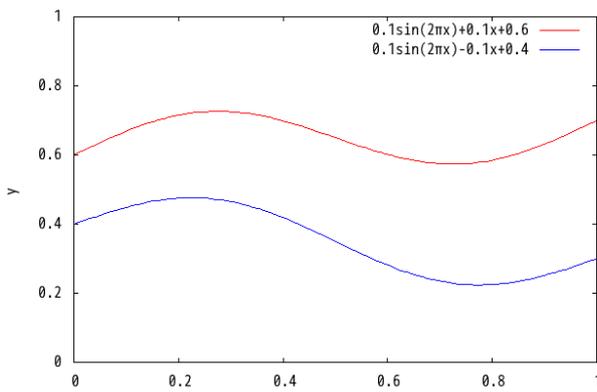


図 5: 本実験において INN が近似する目標区間関数(2)

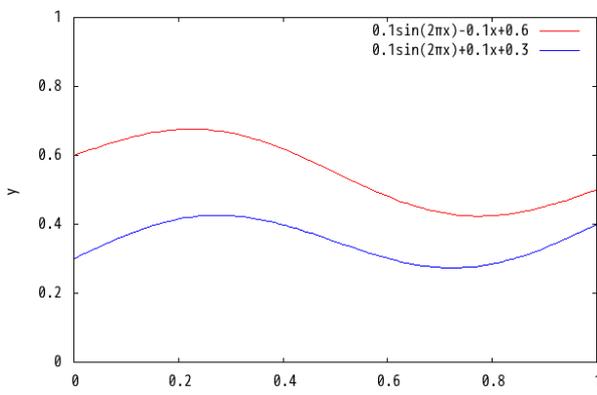


図 6: 本実験において INN が近似する目標区間関数(3)

例題(1)の実験結果を図 7-10 に示す. 図 7・8 (図 9・10) は CW (LU) モデルを用いた場合の結果である. INN および区間 ES を表 1 の通りに設定したとき, 近似誤差最小の解が得られた. 図 7・9 は, 各世代の最良個体による誤差が世代交代に伴って減少する様子を表している. また図 8・10 は, 目標関数, 初期世代および最終世代の最良個体による関数, および, 10 個体 * 10,000 世代 = 全 100,000 個体中の最良個体による関数をそれぞれ表している.

表 1: INN および区間 ES の設定 (例題(1))

	LU モデル	CW モデル
中間層ユニット数	10	3
ES 戦略	(3,10)-ES	(3,10)-ES
世代数	10,000	10,000
式(9)-(12)の rand	[-0.5,0.5]	[-0.5,0.5]
	一様乱数	一様乱数

次に, 例題(2)の実験結果を例題(1)と同様に図 11-14 に示す. 図 11・12 (図 13・14) は CW (LU) モデルを用いた場合の結果である. INN および区間 ES を表 2 の通りに設定したとき, 近似誤差最小の解が得られた.

表 2: INN および区間 ES の設定 (例題(2))

	LU モデル	CW モデル
中間層ユニット数	5	5
ES 戦略	(10,100)-ES	(3,10)-ES
世代数	1,000	10,000
式(9)-(12)の rand	[-0.5,0.5]	[-1.0,1.0]
	一様乱数	一様乱数

最後に, 例題(3)の実験結果を例題(1)(2)と同様に図 15-18 に示す. 図 15・16 (図 17・18) は CW (LU) モデルを用いた場合の結果である. INN および区間 ES を表 3 の通りに設定したとき, 近似誤差最小の解が得られた.

表 3: INN および区間 ES の設定 (例題(3))

	LU モデル	CW モデル
中間層ユニット数	3	5
ES 戦略	(3,10)-ES	(3,10)-ES
世代数	10,000	10,000
式(9)-(12)の rand	[-1.0,1.0]	[-1.0,1.0]
	一様乱数	一様乱数

図 8・10・12・14・16・18 から, 教師なし学習にも関わらず, 最良個体の関数は目標関数を概ねよい精度で近似できていることがわかる.

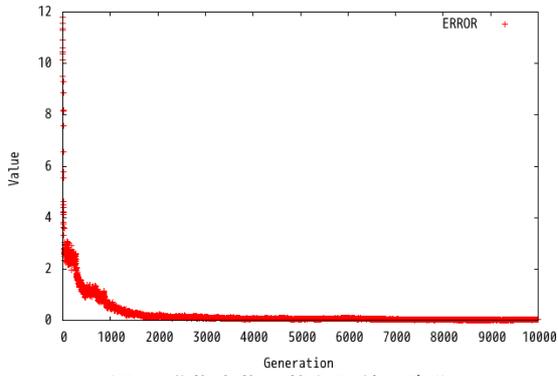


図 7: 世代交代に伴う誤差の変化

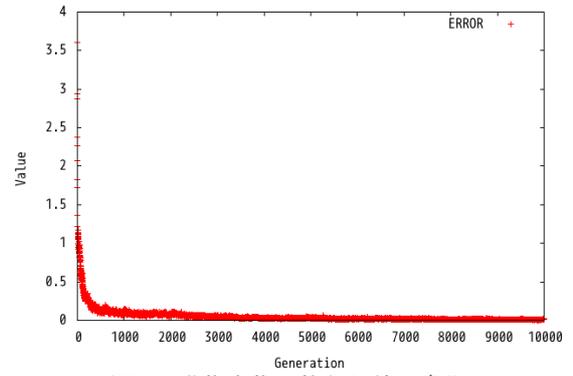


図 11: 世代交代に伴う誤差の変化

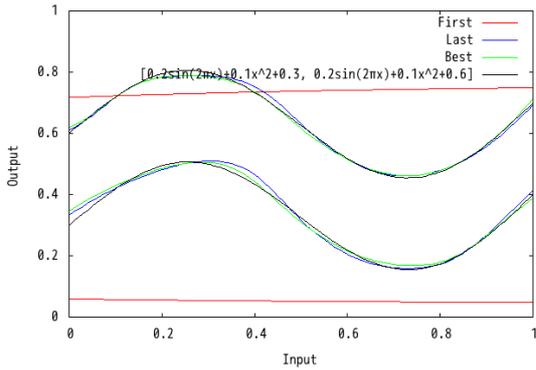


図 8: 目標関数と初期/最終/最良個体による関数

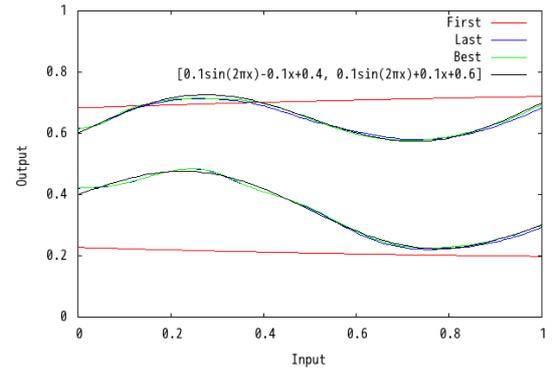


図 12: 目標関数と初期/最終/最良個体による関数

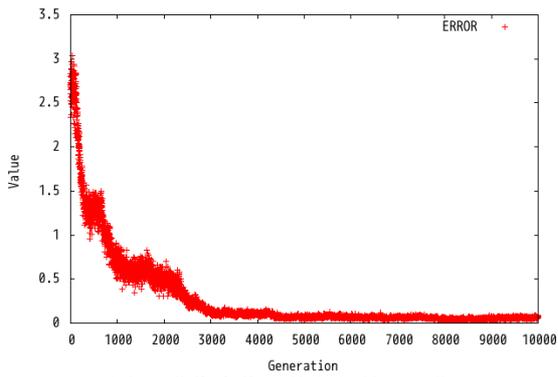


図 9: 世代交代に伴う誤差の変化

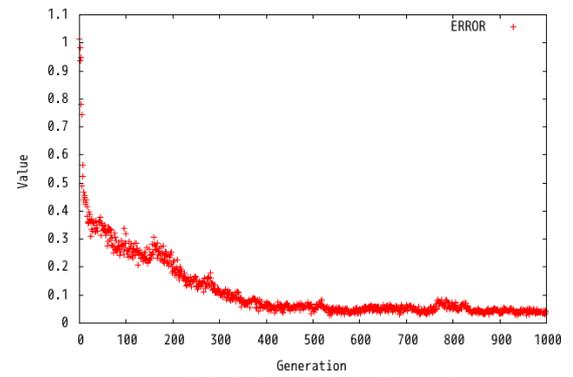


図 13: 世代交代に伴う誤差の変化

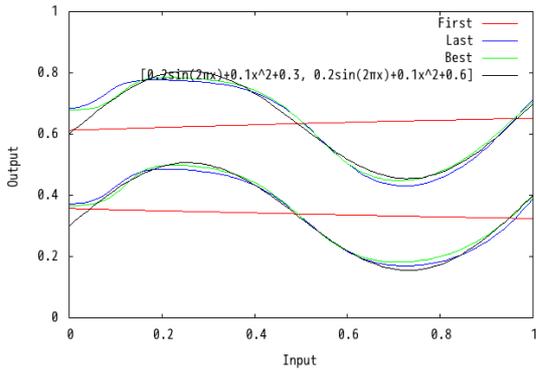


図 10: 目標関数と初期/最終/最良個体による関数

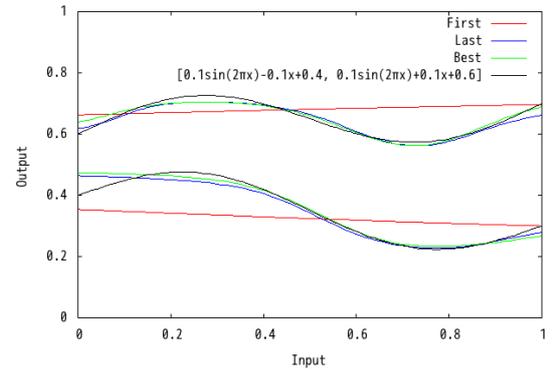


図 14: 目標関数と初期/最終/最良個体による関数

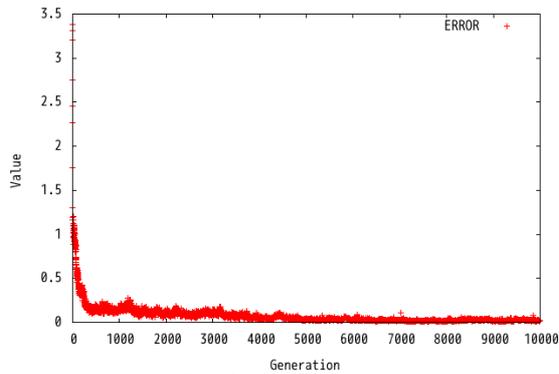


図 15: 世代交代に伴う誤差の変化

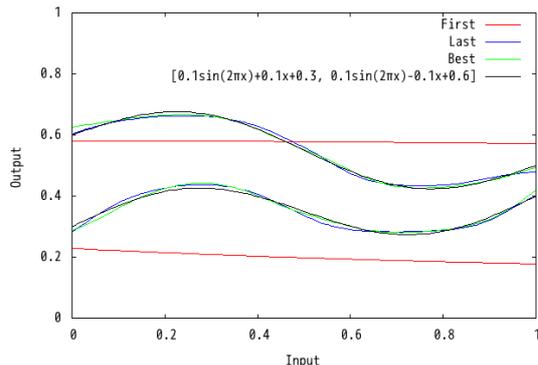


図 16: 目標関数と初期/最終/最良個体による関数

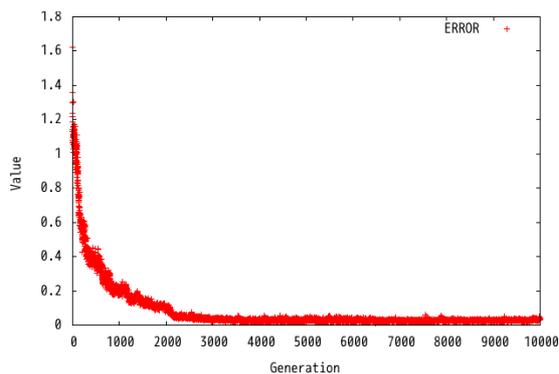


図 17: 世代交代に伴う誤差の変化

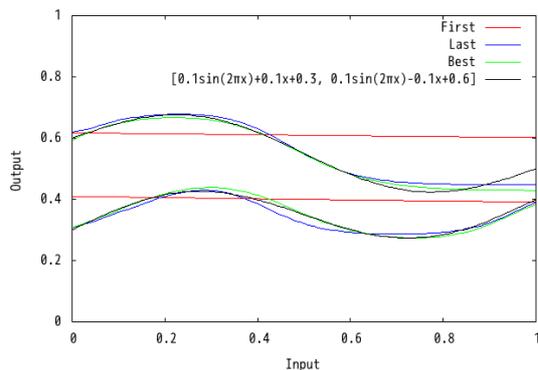


図 18: 目標関数と初期/最終/最良個体による関数

5. 実験結果の考察

INN のパラメータ (中間層ユニット数), 区間 ES のパラメータ (全個体数統一のもとでの集団サイズと世代数など), 目標関数を様々に変えて実験を進めている. これまでの実験の結果から, 以下のような知見が得られてきている.

- LU モデルより CW モデルの方が, 解の進化が良好である.
- $(\mu + \lambda)$ -ES より (μ, λ) -ES の方が近似誤差の小さい解を得やすい.
- (μ, λ) -ES において, λ を μ より相対的に大きくとる方が, 解の進化が良好である.
- 生成個体数 100,000 の条件のもとで, 個体数を相対的に多くした場合より, 世代数を相対的に多くした方が, 解の進化が良好である.

6. まとめ

本稿では, 本研究で提案した区間 ES を用いて INN の教師なし学習を行う方法を示した. 本稿における INN の学習目標は, 明示的には与えられない目標区間関数を近似することである. 実験の結果, 提案手法を用いて INN の区間結合強度と区間しきい値をチューニングすることで, 教師なし学習にも関わらず, 良好な精度で INN に目標関数を近似させられることがわかった. また, 様々な設定のもとでの実験の結果, LU モデルより CW モデルの方が解の進化が良好である. ことなどがわかった.

今後の課題として, 提案手法に関する 5 章の知見の一般性をさらなる評価実験によって確かめてゆくこと, 本研究で提案している他の区間進化計算手法 (区間 GA, 区間 DE, 区間 PSO など[3]) との比較, 他の問題への提案手法の適用に基づいた有効性評価などが挙げられる. 区間 GA に関するこれまでの研究成果は文献[6]を参照されたい.

参考文献

- [1] H. Ishibuchi, H. Tanaka, and H. Okada, An architecture of neural networks with interval weights and its application to fuzzy regression analysis, *Fuzzy Sets and Systems*, 57(1), 27-39, 1993.
- [2] X. Yao, Evolving artificial neural networks, *Proceedings of the IEEE*, 87(9), 1423-1447, 1999.
- [3] 岡田, 解の遺伝子表現に区間値やファジィ値を用いた区間・ファジィ進化計算手法の提案, 情報処理学会関西支部平成24年度支部大会, 2012.
- [4] G. Alefeld and J. Herzberger, *Introduction to Interval Computation*, Academic Press, 1983.
- [5] L.J. Eshelman and J.D. Schaffer, Real-coded genetic algorithms and interval-schemata, in D.L. Whitley (ed), *Foundation of Genetic Algorithms 2*, 187-202, 1993.
- [6] 松瀬, 和田, 山下, 岡田, 区間値遺伝子を用いた区間 GA によるニューロエボリューション, 情報処理学会関西支部平成24年度支部大会, 2012.