進化型多目的最適化アルゴリズムの動的バイナリーニューラルネットワークへの応用

著者	外川 智之
出版者	法政大学大学院理工学・工学研究科
雑誌名	法政大学大学院紀要. 理工学・工学研究科編
巻	62
ページ	1-5
発行年	2021-03-24
URL	http://doi.org/10.15002/00023938

進化型多目的最適化アルゴリズムの動的バイナリーニューラルネットワークへの応用

APPLICATION OF EVOLUTIONARY MULTI-OBJECTIVE OPTIMIZATION ALGORITHM TO DYNAMIC BINARY NEURAL NETWORKS

外川智之
Tomoyuki TOGAWA
指導教員　斎藤利通

法政大学大学院理工学研究科電気電子工学専攻修士課程

This paper studies application of an evolutionary multi-objective optimization algorithm. A dynamic binary neural networks is characterized by the signum activation function and ternary connection parameters. Depending on the connection parameters, this network can generate various binary periodic orbits. In order to evaluate the performance, we consider the bi-objective problem corresponding to stability of the binary periodic orbits and sparsity of the connection parameters. Although uni-objective optimization problems require the optimization of only one objective, multi-objective optimization problems require the simultaneous optimization of multiple objectives. In order to optimize the bi-objective problem, we present a multiobjective evolutionary algorithm based on decomposition. This algorithm decomposes the bi-objective problem into multiple subproblems and can optimize the problem effectively. Performing elementary numerical experiments for typical examples of binary periodic orbits, it is confirmed that the algorithm realizes both strong orbit stability and appropriate connection sparsity. It is also confirmed that the algorithm outperforms another algorithm based on the Lasso regularization.
***Key Words** : Multi-objective evolutionary algorithm, MOEA/D, Binary neural networks.*

１．はじめに

本研究では，進化型多目的最適化アルゴリズムの動的バイナリーニューラルネットワーク(DBNN)への応用について考える．DBNN はシグナム活性化関数，3 値結合パラメータ，整数しきい値によって特徴づけられるリカレントニューラルネットワークである[1]-[5]．結合パラメータによって様々な 2 値周期軌道を呈する．DBNN は厳密な解析とハードウェア実装に適している[4]-[8]．ダイナミクスは点の集合で定義されるデジタルリターンマップ(Dmap)に統合される．DBNN は様々な工学的応用があり，連想メモリ[9]，スイッチング制御[10]，セントラルパターンジェネレータ[11]を含む．

DBNN の解析において，所望の周期軌道(TBPO)の埋め込みとその安定性が基礎的な問題となる．従来の研究[1]より，TBPO の埋め込み条件は分かっており，適切な結合パラメータのスパースは安定性を強めることも示唆されている．結合パラメータのスパースと TBPO の安定性の 2 目的問題の最適化は工学的応用において，低電力かつロバスト性を伴ったシステムを実現させる．しかし，この 2 目的問題にはトレードオフが存在し，最適化することは困難である．

単目的最適化問題は 1 つの目的のみを最適化するが，多目的最適化問題は複数の目的を同時に最適化する必要がある[12]-[15]．本研究では 2 目的問題を最適化するために，MOEA/D[12]を導入する．MOEA/D は複数目的問題を任意の単目的問題に分割できる．2 目的問題を最適化することができれば，結合パラメータに対応するパレートセットを獲得する．

次に，DBNN に関する 2 目的関数を導入する．1 つ目は TBPO に直接落ち込む点(DEPP)の数で，TBPO の安定性に関連する．2 つ目は結合パラメータの 0 の数で，スパースに関連する．セントラルパターンジェネレータを含む TBPO の典型例に数値実験を行うことで，MOEA/D が適切なスパースした結合パラメータを獲得し，TBPO の安定性を強めることを確認する．また，MOEA/D とラッソ正則化[16][17]を比較し優位性を検証する．

２．動的バイナリーニューラルネットワーク

DBNN と TBPO の定義をする．ダイナミクスは次に定義される．

$$
\begin{gathered}
x_i^{t+1} = F\left(\sum_{j=1}^{N} w_{ij} x_j^t - T_i\right), \quad i = 1 \sim N \\
F(x) = \begin{cases} +1 & \text{if } x \geq 0 \\ -1 & \text{if } x < 0 \end{cases}
\end{gathered} \tag{1}
$$

ここで $x_i^t \in \{-1,+1\} \equiv \boldsymbol{B}$ は離散時刻 t における i 番目の 2 値状態ベクトルである．DBNN はシグナム活性化関数 F と 3 値結合パラメータ $w_{ij} \in \{-1,0,+1\}$ と整数しきい値 $T_i \in \{0,\pm1,\pm2,\cdots\}$ によって特徴づけられる．簡単のため，これらのパラメータを行列で表す．

$$
\boldsymbol{W} \equiv \begin{pmatrix} w_{11} & \cdots & w_{1N} \\ \vdots & \ddots & \vdots \\ w_{N1} & \cdots & w_{NN} \end{pmatrix} \quad \boldsymbol{T} \equiv \begin{pmatrix} T_1 \\ \vdots \\ T_N \end{pmatrix} \tag{2}
$$

式 (1) を $\boldsymbol{x}^{t+1} = \boldsymbol{F}_D(\boldsymbol{x}^t)$ と略記し，$\boldsymbol{x}^t \equiv (x_1^t, \cdots, x_N^t)$ である．周期 p の TBPO は 2 値ベクトルの系列である．

$$
\boldsymbol{z}^1, \cdots, \boldsymbol{z}^p, \cdots, \begin{cases} \boldsymbol{z}^t = \boldsymbol{z}^s & \text{for } |t-s| = n \cdot p \\ \boldsymbol{z}^t \neq \boldsymbol{z}^s & \text{for } |t-s| \neq n \cdot p \end{cases} \tag{3}
$$

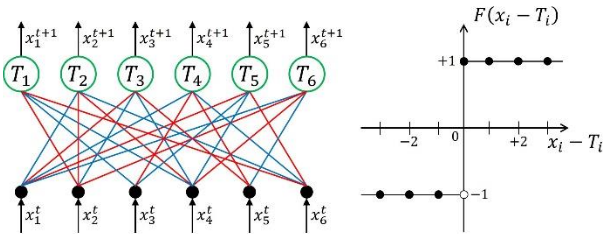

図 1　DBNN とシグナム活性化関数

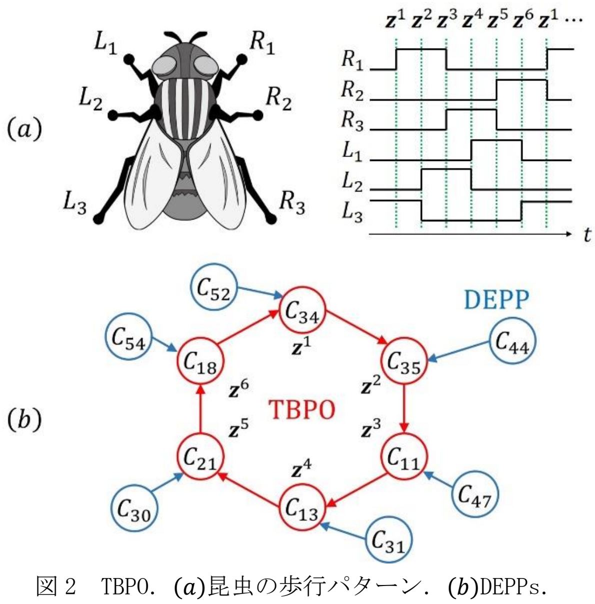

図 2　TBPO．(a)昆虫の歩行パターン．(b)DEPPs．

ここで $\boldsymbol{z}^t \equiv (z_1^t, \cdots, z_N^t)$, $z_i^t \in \boldsymbol{B}$ である．全ての $t = 1 \sim p$, $(\boldsymbol{z}^{p+1} = \boldsymbol{z}^1)$ において $\boldsymbol{z}^{t+1} = \boldsymbol{F}_D(\boldsymbol{z}^t)$ ならば，周期 p の TBPO は DBNN に埋め込まれる．以下の条件を満たすとき，TBPO の埋め込みは保証される．

$$
\begin{aligned}
& L(i) < T_i \leq R(i) \ \text{ for all } i \\
& R(i) = \min_{\tau} \left(\sum_{j=1}^{N} w_{ij} z_j^{\tau}\right) \ \text{ for } \tau \text{ such that } z_i^{\tau+1} = +1 \\
& L(i) = \max_{\tau} \left(\sum_{j=1}^{N} w_{ij} z_j^{\tau}\right) \ \text{ for } \tau \text{ such that } z_i^{\tau+1} = -1
\end{aligned} \tag{4}
$$

[1]で提案された相関学習は $L(i) \leq R(i) + 1$ を満たす w_i と T_i を決める．TBPO の典型例として周期 6 の 6 次元の TBPO を導入する．

$$
\begin{aligned}
\boldsymbol{z}^1 &= (+1,-1,-1,-1,-1,+1) \\
\boldsymbol{z}^2 &= (+1,-1,-1,-1,+1,-1) \\
\boldsymbol{z}^3 &= (-1,-1,+1,-1,+1,-1) \\
\boldsymbol{z}^4 &= (-1,-1,+1,+1,-1,-1) \\
\boldsymbol{z}^5 &= (-1,+1,-1,+1,-1,-1) \\
\boldsymbol{z}^6 &= (-1,+1,-1,-1,-1,+1)
\end{aligned} \tag{5}
$$

この TBPO は図 2(a) に示す昆虫の歩行パターンに対応している．相関学習を適用することで，フルバイナリーの結合パラメータとしきい値を得た．

$$
\boldsymbol{W}_f = \begin{pmatrix} +1 & +1 & -1 & -1 & -1 & +1 \\ -1 & +1 & +1 & +1 & -1 & -1 \\ +1 & -1 & +1 & -1 & +1 & -1 \\ -1 & -1 & +1 & +1 & +1 & -1 \\ +1 & -1 & -1 & -1 & +1 & +1 \\ -1 & +1 & -1 & +1 & -1 & +1 \end{pmatrix} \quad \boldsymbol{T}_f = \begin{pmatrix} 2 \\ 2 \\ 2 \\ 2 \\ 2 \\ 2 \end{pmatrix} \tag{6}
$$

$\boldsymbol{B}^N$ は 2^N の点の集合と等しいので，DBNN のダイナミクスは Dmap に統合される．

$$
\theta^{t+1} = f_D(\theta^t), \theta^t \in L_N \equiv \{C_1, \cdots, C_{2^N}\} \tag{7}
$$

ここで $C_i \equiv \frac{i}{2^N}$, $i = 1 \sim 2^N$ である．Dmap の領域は $L_6 = \{C_1, \cdots, C_{2^6}\}$ で，$C_1 = (-1,-1,-1,-1,-1,-1)$，$C_6 = (+1,+1,+1,+1,+1,+1)$ である．マップ上の点の数は有限なので，定常状態は周期軌道となる．

Dmap を使い，周期点の定義をする．$f_D^p(\theta_p) = \theta_p, f_D^k(\theta_p) \neq \theta_p, 1 \leq k < p$ を満たすとき，ある点 $\theta_p \in L_D$ は周期 p の周期点である．周期点の系列 $\{f_D(\theta_p), \cdots, f_D^p(\theta_p)\}$ を周期軌道(BPO)と呼ぶ．周期点でない $\theta_e \in L_D$ が q ステップで周期軌道に落ち込む場合，$\theta_e \in L_D$ を E 周期点(EPP)と呼ぶ．また，1 ステップで周期軌道に落ち込む EPP を直接 E 周期点(DEPP)と呼ぶ．DEPP の数が増すほど，TBPO の安定性が強まる．図 2(b) に TBPO と DEPPs を示す．また，図 3(a) にフルバイナリーの DBNN と Dmap を示す．

３．DBNN の安定性とスパース

TBPO の安定性と結合パラメータのスパースについて考える．DBNN を評価するために，2 目的関数を導入する．1 つ目の目的は周期pの TBPO の安定性である．

$$F_1(\boldsymbol{W}) = 1 - \frac{\#\text{DEPPs falling into a TBPO}}{2^N - p} \qquad (8)$$

$F_1(\boldsymbol{W})$ が小さくなるほど安定性が強まる．2 つ目の目的は結合パラメータのスパースである．

$$F_2(\boldsymbol{W}) = 1 - \frac{\#\text{zeros in connection matrix}}{N^2 - N} \qquad (9)$$

$F_2(\boldsymbol{W})$ が小さくなるほどスパースである．フルバイナリーの結合パラメータ$\boldsymbol{W}_f$では$F_1(\boldsymbol{W}_f) = 52/58, F_2(\boldsymbol{W}_f) = 30/30$である．$\boldsymbol{W}_f$ に適切に 0 を挿入することで，TBPO の安定性が強まる．スパースした結合パラメータの例を次に示す．

$$\boldsymbol{W}_s = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & +1 \\ 0 & 0 & 0 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & +1 & -1 \\ 0 & 0 & +1 & 0 & 0 & -1 \\ 0 & -1 & -1 & 0 & 0 & 0 \\ -1 & +1 & 0 & 0 & 0 & +1 \end{pmatrix} \; \boldsymbol{T}_s = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \qquad (10)$$

これらのパラメータは式(4)の条件を満たしており，埋め込み保証されている．図 3(b)に$\boldsymbol{W}_s$の DBNN と Dmap を示す．結合パラメータ$\boldsymbol{W}_s$の各評価値は$F_1(\boldsymbol{W}_s) = 22/58, F_2(\boldsymbol{W}_s) = 6/30$である．2 目的問題を最適化するために，2 つのアルゴリズムを考える．

（１）MOEA/D

1 つ目のアルゴリズムは MOEA/D であり，次の問題を最小化する．

$$\text{Minimize } \; F(\boldsymbol{W}) = (F_1, F_2) \in S_O, \text{subject to } \; \boldsymbol{W} \in S_D \qquad (11)$$

ここで $S_O = \{(F_1, F_2) | 0 \le F_1 \le 1, 0 \le F_2 \le 1\}$，$S_D = \{\boldsymbol{W} | w_{ij} \in \{-1, 0, +1\}, i = 1 \sim N, j = 1 \sim N\}$ である．S_O, S_Dはそれぞれ目的関数空間とパラメータ空間を表している．次に，支配関係を定義する．以下の条件を満たすとき，$\boldsymbol{W}_1 \in S_D$は$\boldsymbol{W}_2 \in S_D$を支配しているといえる．

$$F_1(\boldsymbol{W}_1) < F_1(\boldsymbol{W}_2) \text{ and } F_2(\boldsymbol{W}_1) < F_2(\boldsymbol{W}_2) \qquad (12)$$
$$F_1(\boldsymbol{W}_1) < F_1(\boldsymbol{W}_2) \text{ and } F_2(\boldsymbol{W}_1) = F_2(\boldsymbol{W}_2) \qquad (13)$$
$$F_2(\boldsymbol{W}_1) < F_2(\boldsymbol{W}_2) \text{ and } F_1(\boldsymbol{W}_1) = F_1(\boldsymbol{W}_2) \qquad (14)$$

どのパラメータにも支配されない$\boldsymbol{W}_p$が存在する場合，これをパレート解と呼ぶ．パレート解の集合をパレートセットと呼び，目的関数空間上にパレートフロントを形成する．

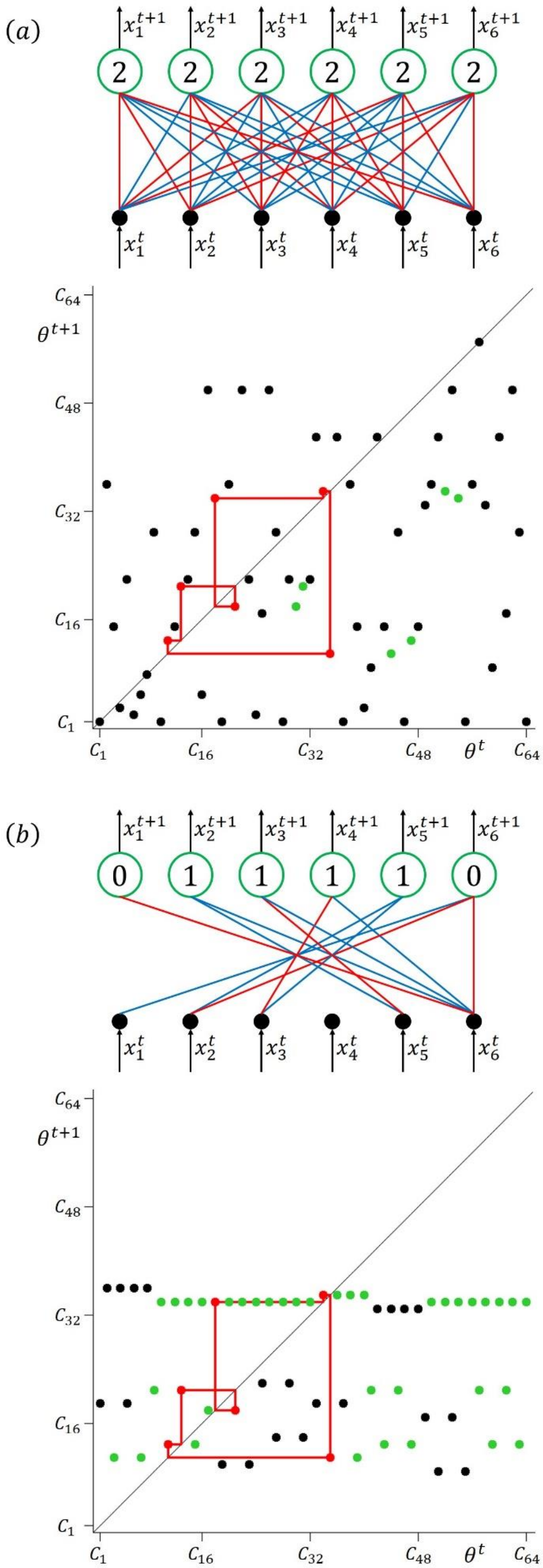

図 3　DBNN と Dmap. (a)フルバイナリーの結合パラメータ$\boldsymbol{W}_f(f_1(\boldsymbol{W}_f) = 52/58, f_2(\boldsymbol{W}_f) = 30/30)$. (b)スパースの結合パラメータ$\boldsymbol{W}_s(f_1(\boldsymbol{W}_s) = 22/58, f_2(\boldsymbol{W}_s) = 6/30)$.

MOEA/D は weighted sum を用い，2 目的問題を任意の単目的問題に分割する．

$$\text{Min } F^k(\boldsymbol{W}|\boldsymbol{\lambda}^k) = \lambda_1^k F_1(\boldsymbol{W}) + \lambda_2^k F_2(\boldsymbol{W}), k = 1 \sim M_1 \quad (15)$$

ここで重みベクトルは $\boldsymbol{\lambda}^k \equiv (\lambda_1^k, \lambda_2^k), k \in \{1, \cdots, M_1\}$ で，$\lambda_1^k = (k-1)/(M_1-1), \lambda_1^k + \lambda_2^k = 1$ によって与えられる．各重みベクトルが 1 つの単目的問題に対応する．個体群 $\{\boldsymbol{W}^1, \cdots, \boldsymbol{W}^{M_1}\}$ はM_1個の個体から成り，$\boldsymbol{W}^k$ は結合パラメータに対応するk番目の個体である．外部個体群(EP)は各世代において非劣解を保存する．$\boldsymbol{\lambda}^{l_1}, \boldsymbol{\lambda}^{l_2}(l_1, l_2 \in B(k))$ は $\boldsymbol{\lambda}^k$の近傍ベクトルであり，$B(k) = \{k-1, k, k+1\}$ である．アルゴリズムを次に定義する．

Step 1（初期化）

$\mathrm{EP} = \emptyset$ とし，初期個体を生成する．

Step 2（更新）

近傍から個体を 1 つランダムに選択する．結合パラメータにランダムに 0 を 1 つ挿入する突然変異を適用し，子個体生成する．生成した子個体からしきい値を決める．エリート選択により近傍解を更新する．子個体に支配された EP を除去し，非劣解となる子個体を EP に追加する．

Step 3（終了条件）

最終世代に至るまで Step 2 を繰り返す．アルゴリズム終了後，EP がパレート解を与える．

アルゴリズムの各パラメータは$M_1 = 50, g_{max} = 30$を適用した．Step 1 で生成した個体はフルバイナリーであり，0 を 1 つ挿入することで初期個体が与えられる．図 4 に進化過程を示す．MOEA/D を適用することで，3 つのパレート解を獲得した．その中の 1 つは式(10)に示した$\boldsymbol{W}_s$である．このパレートセットは強い安定性と適切なスパースを実現する．

（２）ラッソ正則化

2 つ目のアルゴリズムはラッソ正則化であり，次の問題を最小化する．

$$\text{Min } G(\boldsymbol{W}|\boldsymbol{\lambda}) = \lambda_1 F_1(\boldsymbol{W}) + \lambda_2 F_2(\boldsymbol{W}), \lambda_1 + \lambda_2 = 1 \quad (16)$$

ラッソ正則化は機械学習の正則化の 1 つである．正則化において，コスト関数はエラー項と L1 ノルムに重み付けされた形で与えられる．F_1はエラーに対応しており，F_2は L1 ノルムに対応している．ノルム項を加えることでオーバーフィッティングを効果的に防ぐ．アルゴリズムを次に定義する．

Step 1（初期化）

初期個体を生成する．

Step 2（更新）

結合パラメータにランダムに 0 を 1 つ挿入する突然変異を適用し，子個体生成する．生成した子個体からしきい値を決める．$G(\boldsymbol{W}|\boldsymbol{\lambda})$により評価された子個体をエリート選択により選択する．

Step 3（終了条件）

最終世代に至るまで Step 2 を繰り返す．

アルゴリズムの各パラメータは$\lambda_1 = \lambda_2 = 0.5, M_2 = 200, g_{max} = 30$を適用した．図 4 に進化過程を示す．ラッソ正則化を適用することで，1 つの最適解を獲得した．

（３）２つのアルゴリズムの比較

図 5 に示す 2 つのアルゴリズムの比較より，5 世代目以降，MOEA/D の評価値Fがラッソ正則化の評価値Gより小さいことがわかる．これらの結果は MOEA/D がラッソ正則化より優れていることを示唆している．また，このような結果は MOEA/D の探索能力と効果的な個体選択によるものだと考えられる．MOEA/D は複数の重みベクトルを使い目的関数空間上を幅広く探索できる．また，近傍解を使うことで子個体生成の豊富さがある．一方，ラッソ正則化は重みベクトルが固定されているため，各個体が制限されている．

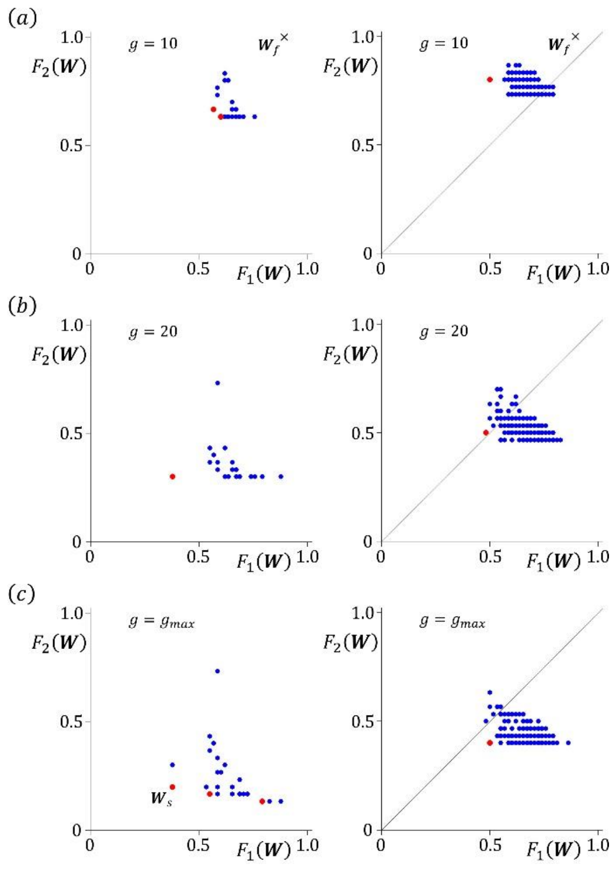

図 4　MOEA/D(左)とラッソ正則化(右)の進化過程．

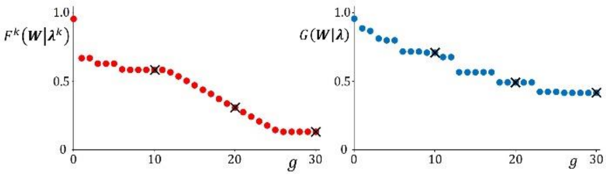

図 5　MOEA/D(左)とラッソ正則化(右)の比較.

アルゴリズムの性能を確認するため，周期 6 の TBPO を 10 個準備し数値実験を行った．これらの TBPO はランダムに生成したものであり，次の 3 つのパラメータを適用した：MOEA/D($M_1 = 50$)，ラッソ正則化($M_2 = 50, M_2 = 200$)．初期個体はフルバイナリーの結合パラメータより生成し，世代上限$g_{max} = 30$である．10 個の TBPO の評価値を表 1 に示す．表 1 では，MOEA/D がラッソ正則化の最良値より良い結果を示している．

表 1　周期 6 の TBPO10 個の評価値

Generation		10	20	30
MOEA/D	Avg	0.523	0.308	0.125
ラッソ正則化 $M_2 = 50$	Best	0.545	0.333	0.129
	Avg	0.647	0.476	0.381
	Worst	0.671	0.610	0.597
ラッソ正則化 $M_2 = 200$	Best	0.536	0.332	0.129
	Avg	0.645	0.476	0.379
	Worst	0.710	0.616	0.601

以上のことから，MOEA/D は DBNN の学習において TBPO の安定性の強化と適切なスパース結合に有用であると考えられる．

4．むすび

本研究では TBPO の安定性と DBNN のスパースについて考えた．2 目的問題を最適化するために，MOEA/D を導入した．典型的な TBPO に数値実験を行うことで，MOEA/D が強い安定性と適切なスパースを満たした最適解を獲得できることを確認した．また，従来の学習方法であるラッソ正則化と比較をすることでその優位性を示した．今後の課題として，様々な TBPO の数値実験，他の目的への応用，進化過程の解析，ハードウェア実装などが挙げられる．

参考文献

1) R. Sato and T. Saito, Stabilization of desired periodic orbits in dynamic binary neural networks, Neurocomputing 248, pp. 19-27, 2017.

2) R. Sato, S. Aoki, and T. Saito, Connection sparsity versus orbit stability in dynamic binary neural networks, Proc. IJCNN, pp. 4482-4487, 2017.

3) T. Togawa and T. Saito, Connection Sparsification and Orbit Stabilization of Dynamic Binary Neural Networks based on Multiobjective Evolutionary Algorithms, Proc. IEEE/WCCI, DOI: 10.1109/IJCNN48605.2020.9206599, 2020.

4) S. Aoki, S. Koyama, and T. Saito, Theoretical analysis of dynamic binary neural networks with simple sparse connection, Neurocomputing, 341, pp. 149-155, 2019.

5) S. Anzai, S. Koyama, S. Aoki, and T. Saito, Sparse Dynamic Binary Neural Networks for Storage and Switching of Binary Periodic Orbits, T. Gedeon et al. (Eds.): ICONIP 2019, LNCS 11954, pp. 536-542, 2019.

6) D. L. Gray and A. N. Michel, A training algorithm for binary feed forward neural networks, IEEE Trans. Neural Netw. 3, 2, 176-194, 1992.

7) M. Courbariaux, Y. Bengio, and J.–P. David, BinaryConnect: Training Deep Neural Networks with binary weights during propagations, in NIPS, pp. 3105-3113, 2015.

8) M. Courbariaux, I. Hubara, D. Soudry, R. EI-Yaniv, and Y. Bengio, Binarized Neural Networks: Training Neural Networks with Weights and Activations Constrained to +1 or -1, arXiv: 1602.02830, 2016.

9) J. J. Hopfield, Neural networks and physical systems with emergent collective computation abilities, Proc. Of the Nat. Acad. Sci., 79, pp. 2554-2558, 1982.

10) B. K. Bose, Neural Network Applications in Power Electronics and Motor Drives-An Introduction and Perspective, IEEE Trans. Ind. Electron. 54, 1, pp. 14-33, 2007.

11) A. Lozano, M. Rodriguez, and R. Roberto Barrio, Control strategies of 3-cell Central Pattern Generator via global stimuli, Sci. Rep. 6:23622, 2016.

12) Q. Zhang and L. Hui, MOEA/D: A Multiobjective Evolutionary Algorithm Based on Decomposition, IEEE Trans. Evol. Comput., 11, 6, pp. 712-731, 2007.

13) K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, A Fast and Elitist Multiobjective Genetic Algorithm: NSGA-II, IEEE Trans. Evol. Comput. 6, 2, pp. 182-197, 2002.

14) R. Wang, Z. Zhou, H. Ishibuchi, T. Liao, and T. Zhang, Localized Weighted Sum Method for Many-Objective Optimization, IEEE Trans. Evol. Comput., 11, 1, pp. 3-18, 2018.

15) C. He, L. Li, Y. Tian, X. Zhang, R. Cheng, Y. Jin, and X. Yao, Accelerating Large-Scale Multiobjective Optimization via Problem Reformulation, IEEE Trans. Evol. Comput., 23, 6, pp. 949-961, 2019.

16) R. Tibshirani, Regression Shrinkage and Selection via the Lasso, J. Roy. Stat. Soc. B, 58, 1, pp. 267-288, 1996.

17) B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, Least Angle Regression, Ann. Statist., 32, 2, pp. 407-499, 2004.