

タブー探索を用いた拡張型 MCRNN の構造最適化

日大生産工 ○二宮 力也 日大生産工 山内 ゆかり

1 まえがき

リカレントニューラルネットワークとは、その構造およびユニット間の接続重みのパラメータを適切に設定することで時系列データに対して高い予測精度の予測が可能であるが、多数ユニットから構成されるリカレントニューラルネットワークの構造を最適に設定することは難しい。また、ニューラルネットワークの予測能力は学習用データに対する誤差の最小化だけではなく、学習に使用されない未知のデータに対する汎化能力が必要である。

ニューラルネットワークは構成するユニット数が少ない場合には適切に学習を行うことが難しく、逆にユニット数が多い場合には過学習による汎化能力の低下が起こる。このため、ネットワークを構成するユニットの数などのネットワーク構造を設計者が適切に設定する必要があり、設計者が試行錯誤的にネットワークの構造を決定していた。

ニューラルネットワークが高精度の近似、あるいは予測能力を獲得するためには、構成するユニット間の接続重み、学習ルール、ネットワーク構造を最適に設定する必要がある。しかし、対象問題に対応した高い精度の予測能力を持つネットワークの構造を決定するためには、多くの調整時間が必要であるため、遺伝的進化計算手法によりネットワークの適切な構造を発見する手法[1]や、接続重みを含めたネットワーク構造を遺伝的進化手法により学習するTWEANNを基礎とした構造最適化手法[2]、ユニット間の伝達関数の最適化手法[3]などが提案されている。この他にも、ニューラルネットワークの最適化の関連研究が多く報告されている[4]。

林田ら[5]は有効であると考えられる解の近傍を集中的に探索できるタブー探索を用いてMCRNN(Multi-Context-RNN)の構造決定する方法により予測能力において有効性を示した。だ

が、未知のデータに対する汎化能力では過去の結果より少し劣る結果になった。

Katagiri et al[6]が提案した、拡張型MCRNNにエリート保存、自己適応型突然変異などのGAの手法を取り入れた最適化手法は、汎化能力において有効性を示していた。

そこで本研究では、タブー探索を用いて拡張型MCRNNの構造を作成し、学習用データに対する予測能力と未知のデータに対する汎化能力の両方に優れている構造の決定を目指す。

2 従来手法

林田らはタブー探索を用いて MCRNN の構造を最適化する手法を提案した。タブー探索は局所的な近傍探索に優れた性能を発揮するので、類似した構造を持つニューラルネットワークのパフォーマンスの改善に有効であることに着目した。

MCRNN の構造は、入力層、中間層のユニット数、およびフィードバック層数により特徴づけられる。ユニット間の接続を適切に選択することで MCRNN の汎化能力の向上が期待される。図1に従来研究に用いられる MCRNN の構造を示す。

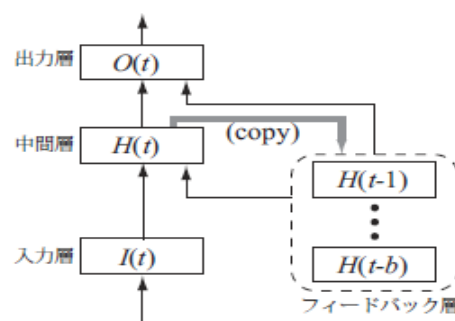


図1 MCRNN

第1段階では MCRNN を構成するユニット数を決定し、第2段階ではユニット間接続の選択を行う。具体的には、MCRNN に対して、入

Structural Optimization of The Expansion Multi Context Neural Networks
Using Tabu Search

－Rikiya NINOMIYA ,Yukari YAMAUCHI

力層、中間層のユニット数、およびフィードバック層数の最適化にタブー探索を適用したあと、さらにユニット間接続の有無の最適化に関してもタブー探索を適用して、任意の時系列データに対する高い予測精度を持つように MCRNN の構造を最適化する。

3 提案手法

本研究では、図 2 に示される拡張型 MCRNN の構造を最適化する手法を提案する。

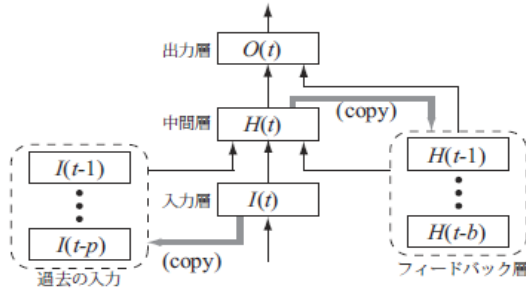


図 2 拡張型 MCRNN

拡張型 MCRNN の入力層、中間層のユニット数、およびフィードバック層数、過去の入力の期数など全体の構造をタブー探索により最適化することで、学習データに適切な学習を行う。さらにユニット間接続の有無の最適化に関してもタブー探索を適用し汎化能力の向上につながると考えられる。2 段階の最適化を行うことで、任意の時系列データに対する高い予測精度を持つ、拡張型 MCRNN の実現を目指す。

4 実験環境

従来の MCRNN では過去の数期分の中間層からの出力を保持するための複数のフィードバック層を持っていたが、提案手法で用いる拡張型 MCRNN は、過去の数期分の入力層からの出力も保持している。

入力層ユニット数 n 、中間層ユニット数 h 、過去の入力の期数 p 、フィードバック層数 b で特徴づけられる解 $A^u = (n, h, p, b)$ としてタブー探索を行い、接続重みを学習した後の近傍解を評価する。

タブー探索では短期メモリおよび長期メモリを用いた探索の集中と多様化を実現している。 g 回目の探索における解を A_g^u として、 A_g^u の各要素のうちの 1 つを 1 だけ増加もしくは、減少させた解の集合

$$\{(n-1, h, p, b), (n+1, h, p, b), (n, h-1, p, b), (n, h+1, p, b), (n, h, p-1, b), (n, h, p+1, b), (n, h, p, b-1), (n, h, p, b+1)\}$$

に対して、これから短期メモリに記録された解を除いたものを A_g^u の近傍集合 AN_g^u とする。探索の集中化のために行われる近傍探索では、現在の解 A^u の近傍のうちもっとも評価の高い解に移動し、これを次期の解とする。すでに探索した解を一定期間短期メモリに記録し、この短期メモリに記録された解への移動を禁止することで、探索の巡回を防ぐ。

短期メモリを用いた近傍探索だけでは、一部の解領域のみの探索となる場合がある。このため、探索領域をいくつかの部分領域に分割して各部分領域における探索回数を保持する長期メモリを用いて、探索の多様化を行う。

d_n, d_h, d_b, d_p をそれぞれ入力層ユニット数 n 、中間層ユニット数 h 、フィードバック層数 b 、過去の入力の期数 p に対応する領域分割数、 w_n, w_h, w_b, w_p を対応する分割幅としたとき、部分領域は、

$$F_{ijkl} =$$

$$\left\{ \begin{array}{l} (n_i, h_j, b_k, p_l); \\ (i-1)w_n < n_i \leq iw_n, (j-1)w_h < h_j \leq jw_h \\ (k-1)w_b < b_k \leq kw_b, (l-1)w_p < p_l \leq lw_p \end{array} \right\}$$

$$i = 1, \dots, d_n, j = 1, \dots, d_h$$

$$k = 1, \dots, d_b, l = 1, \dots, d_p$$

と表される。各部分領域において近傍探索を繰り返す集中的な探索が終われば、探索の多様化のため確率的に別の部分領域に移動し、移動した小領域に含まれるように各パラメータがランダムに決定された初期解が新たに生成され探索が再開される。ここで、 L_{ijkl} は部分領域 F_{ijkl} に移動する確率 p_{ijkl} は次式で与えられる。

$$P_{ijkl} = \frac{L^{range} - L_{ijkl}}{\sum_{i=1}^{d_n} \sum_{j=1}^{d_h} \sum_{k=1}^{d_b} \sum_{l=1}^{d_p} (L^{range} - L_{ijkl})} \quad (1)$$

ただし、 $L^{range} = \max L_{ijkl} - \min L_{ijkl}$ である。

式(1)より、これまでの探索回数が少ない部分領域が高い確率で選択されやすくなり、広範囲の探索を可能にしている。

学習過程におけるニューラルネットワークの評価にはネットワークの出力と学習用データとの誤差が用いられるが、本研究では、過学習による汎化能力の低下に対応するために、汎化能力も考慮した解の評価を行う。このために、学習用データを学習用と検証用に 2 分割し、学

学習誤差と検証誤差のうち大きいものを最小化することで、学習誤差が小さく汎化能力の高いニューラルネットワークの獲得を目指す。ただし、本研究で取り扱うような時系列データにおいて、ある時刻のデータはそれ以前のデータに依存していることが考えられるため、ある期を境界としてデータの分割を行うことが望ましい。特定の区間のデータのみ過剰に適合することを避けることと、計算量とのトレードオフを考慮してデータを3分割してこの操作を2回行う。学習用のデータを3分割したものを、 D_1, D_2, D_3 とする。ただし、学習用の時系列データの時間的範囲を $[0, T]$ としたとき、事前に決めた期 T_1, T_2 を D_1, D_2, D_3 の境界とする。

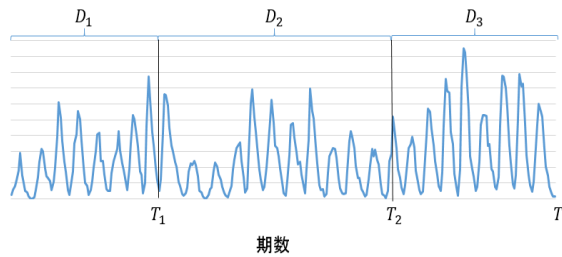


図3 時系列データDの分割

まず、解 A^u により与えられる構造を持つ拡張型 MCRNN を用いて D_1 で誤差逆伝搬法により接続重みを学習し、 D_2 で検証を行う。

$0 \leq \tau_t \leq T_1 < \tau_v \leq T_2$ として学習誤差を e_1^t 、検証誤差を e_1^v とする。次に $D_1 \cup D_2$ で学習、 D_3 で検証を行い、 $0 \leq \tau_t \leq T_2 < \tau_v \leq T$ として、学習誤差を e_2^t 、検証誤差を e_2^v を計算する。ここで対象データを o 次元の時系列データとしたとき、 t 期目における、 s 番目のデータのニューラルネットワークからの出力を $\psi_s(t)$ 、教師信号となる時系列データを $y_s(t)$ として式(2)により誤差を計算する。

$$\begin{aligned} e_i^t &= \sum_{\tau_1} \sum_{s=1}^o (\psi_s(\tau_t) - y_s(\tau_t))^2, \quad i=1,2, \\ e_i^v &= \sum_{\tau_1} \sum_{s=1}^o (\psi_s(\tau_v) - y_s(\tau_v))^2, \quad i=1,2, \end{aligned} \quad (2)$$

解 A^u はこれらの誤差をもちいて、

$$\begin{aligned} E_v(A^u) &= \lambda \max(e_1^t, e_1^v) + (1 - \lambda) \max(e_2^t, e_2^v) \\ \lambda &\in [0, 0.5] \end{aligned} \quad (3)$$

により評価する。本研究では $E_v(A^u)$ を解 A^u に対応する拡張型 MCRNN の加重誤差とよぶ。

一定期間、上記タブー探索により加重誤差の

最良値が更新されない、もしくは一定回数の探索が行われれば、拡張型 MCRNN を構成するユニット数を確定し、ユニット間接続の選択を行う。

ユニット間接続の選択では、探索の効率化のために重要性の低い接続を削除したものを初期解としてタブー探索を行う。得られた拡張型 MCRNN に対して、ユニット間の接続を1本削除し、加重誤差が小さくなる場合、その接続は重要性の低い接続と判断して削除する。順番に接続の削除を行ったものを初期解としてタブー探索を用いてユニット間接続の選択を行う。

2段階のタブー探索を用いた拡張型 MCRNN のユニット数とユニット間接続の決定に基づき、構造最適化を行う。提案手法のアルゴリズムの概要は以下の通りである。

Step 1. (初期化) 初期解 $A_0^u = (n, h, b, p)$ をランダムに生成し、 $A_{best}^u = A_0^u$ とする。 $g = 1$ として、Step2に進む

Step 2. (ユニット数の決定)

Step 2-1. 解 A_{g-1}^u の近傍集合 AN_{g-1}^u のうち加重誤差が最も小さい解を、 A_g^u とする。

Step 2-2. $E_v(A_g^u) < E_v(A_{best}^u)$ が成り立つ場合、 $A_{best}^u = A_g^u$ として A_{best}^u を短期メモリに追加的に記録する。このとき、短期メモリに記録されている解の数が一定数以上の場合、最も古く記録された解を短期メモリから削除する。一定期間最良解の更新が行われず、あるいは一定回数の近傍探索が行われたら短期メモリを更新してStep 2-3 へ。そうでなければ $g = g+1$ としてStep 2-1 へ。

Step 2-3. 最初期化が一定回数行われる。または一定期間最良解の更新がなければ、Step 3 へ。そうでなければ、解の再初期化を実施する。すなわち、長期メモリから定義される式(1)を用いて現在の解とは異なる部分領域で新たな初期解を生成し、長期メモリを更新してstep 2-1へ。

Step 3. (ユニット間接続の選択)

Step 3-1. 現在の最良解 A_{best}^u に対応する拡張型MCRNNに含まれる接続のうち、削除することで加重誤差 E_v が小さくなる接続を削除したものを、接続選択の初期解 A_0^l とし、 $A_{best}^l = A_0^l$ とする。 $s = 1$ としてStep 3-2 へ。

Step 3-2. A_{s-1}^l の近傍集合 AN_{s-1}^l のうち加重誤差が最も小さいものを現在の解 AN_s^l として、短期メモリを更新する。

Step 3-3. $E_v(A_s^l) < E_v(A_{best}^l)$ が成り立つ場合、 $A_{best}^l = A_s^l$ とする。

Step 3-4. Step 3-2、Step3-3 を一定回数繰り返したら、 A_{best}^l を解として終了する。そうでなければ $s = s+1$ としてStep 3-2へ。

提案手法ではStep 2 でユニット数の決定、Step 3 でユニット間の接続がそれぞれ行われる。

5 実験結果

提案手法を適用する時系列データとして、太陽黒点数データ(SSD)[7]を用いる。

本研究では、時系列データを学習用データと汎化能力測定用データとして2分割し、それぞれ10回ずつのシミュレーション実験を行う。

表1 時系列データの期数

	全期数	学習用 (期数)	汎化能力 測定用(期数)
SSD	310	270	40

表2 実験結果

	従来手法	提案手法
最小	100.96	452.57
最大	184.75	501.87
平均	139.53	471.66

表3 構造最適化した結果

	n	h	b	p
最小	4	2	1	1
最大	8	10	3	11
平均	5.9	5.4	2	3.1

従来研究よりも誤差が大きくなってしまったが、適切な拡張型MCRNNの構造は実現できた。近傍の探索が重点的に行われているため、別の領域のいい解に移動できていないものと思われる。

また、データが1次元の時系列データだったので、あまり複雑なネットワークを構築すると予測精度が高くないのではないかと考えられた。多次元の時系列データにも対応したネットワークを構築することが今後の課題であると考えられる。

6 まとめ

本研究では、タブー探索法を用いて拡張型MCRNNの入力層ユニット数、中間層ユニット数、フィードバック層、過去の入力 of 期数および、各ユニット間の接続を選択する構造最適化手法を提案した。

構造は実現できていたが、学習データに対し適切な学習がされていなかった。

初期解を生成するときランダムで生成していたので、ある程度データによって初期解を設定していくことでさらに構造最適化の効率をあげていけるのではないかと考えられる。

「参考文献」

- [1] X.Cai,Z.Zhang,G.Venayagamoorthy and D.Wunsch:Timeseries prediction with recurrent neural networks using a hybrid PSOEA algorithm, *Proceedings of International Joint Conference on Fuzzy Systems*,Budapest,Hungary,pp.1647-1652,2004.
- [2] 大倉、岩波：進化型人工神経回路網の構造進化のための一手法-二十倒立振子問題への適用-,システム制御情報学会論文誌19,pp.51-58,2006.
- [3] G.Mani:Learning by gradient descent in function space,In *Proceedings of IEEE International Conference, System, Man, and Cybernetics*, Los Angeles, CA, pp.242-247,1990.
- [4] X.Yao:Evolving Artificial Neural Networks,*Proceedings of the IEEE* 87, pp.1423-1447,1999.
- [5] 林田 智弘, 西崎 一郎, 末宗 明恵, タブー探索を用いたリカレントニューラルネットワークの構造最適化, 知能と情報(日本知能情報ファジィ学会誌)Vol.27,No.3, pp.638-649(2015)
- [6] H.Katagiri,I.Nishizaki,T.Hayashida and T.Kadoma:Multiobjective evolutionary optimization of training and topology of recurrent neural networks for time-series prediction, *The computer Journal* 55, pp.325-336.2012
- [7] NOAA satellite and information service,[online].Available: <http://www.ngdc.noaa.gov/ngdc.html>