ハイパフォーマンス
コンピューティング 77－2
（1999．8． 2）

ニューラルネットワークによる時系列現象の予測

上田宝生、朱含希、木幡賢史、青山智夫
宮崎大学 工学部 電気電子工学科

階層型ニューラルネットワークと関数の巡回表現を用いて時系列現象を予測する方法を研究した。この方法は、離散化した関数を小区間に分解し、その部分断片を組み合わせて予測する。断片集合の中に関数の将来の形と同じものが存在すれば高精度で予測できる。同じ断片がない場合、小区間の学習から元関数の近似関数がネットワーク内に構成できれば（この可能性は低くないが）精度良く予測できる。

Extrapolations by using neural-networks and a recurrent representation of functions

Hosei UEDA, Hanxi ZHU, Kenji KOBATA, and Tomoo AOYAMA
The Faculty of Engineering, Miyazaki University
Gakuen Kibanadai-nishi 1-1, Miyazaki 889-2192, Japan
E-mail: aoyama@esl.miyazaki-u.ac.jp

Many studies of the extrapolation for time dependent phenomena have been published. In the studies, there were some methods that are adopted neural networks. Using neural networks, explicit descriptions (functions) are not necessary. Futures of phenomena are predicted from observed data only. It is very useful, but the usefulness is a defect at same time. If explicit functions are unknown, it is hard to calculate exactitude of extrapolations. We often find examples of prediction in published papers, but don't discover examinations using neural networks. We consider that the neural network is one of functions. The function is defined at learning data, and it is discrete essentially. But we use it as continuous, make predictions. It is unreasonable. If the neural network is a function, its differential should be defined. The explicit representation is published. Then, function characters near discrete learning points are evaluated by the differential. The fact makes a discrete function into continuous, at least, an index for predictions is got.

1. はじめに

本研究の目的は，時系列現象を階層型ニューラルネットワークを用いて予測する際の精度を調べる方法を提示することである．

最近，時系列現象の予測では進化論的方法有力視されており[1]，ニューラルネットワークが用いられることは少ないなっている．最近の適用例は少ない．しかし我々が試行したところ，進化論的方法に比べニューラルネットによる予測法は時間変化に対し急峻に変化する現象を除けば，ほとんど同一の予測精度であった．[2]また予測ではその精度を示す指標が必要である．それは進化論的予測，ニューラルネットによる予測ともに発表されていないと思われる．

2. 予測理論

以下，関数は一価実関数である．測定時刻間隔が等しい時系列データ{v}の部分集合，

$P_n \equiv \{v_{n-i}, \ldots, v_{n-1}, v_n\}$

をパターンという．関数f(x)の周期を以下のように定義する．

$G(\theta, \eta) = \int_\Omega \{f(x+\theta) - f(x+\eta)\}^2 dx$

このような関数$G(\theta, \eta)$を導入する．Ωは積分区間である．f(x)が周期関数の場合，適切な積分区間Ωをとると，G=0を満足する跡は等間隔な直線である．η'=定数としたとき，$G(\theta, \eta')$値の変化は次の関係を満足する．

① G=0となるθ値 $\{\theta_1, \theta_2, \ldots, \theta_n\}$ は $\theta_2 - \theta_1 = \theta_4 - \theta_3 = \ldots\ldots$ である．

② $\theta_i \leqq \theta \leqq \theta_{i+1}$, $\theta_{i+1} \leqq \theta' \leqq \theta_{i+2}$ である中間のθ, θ'値に対して，

$G(\theta, \eta') = G(\theta', \eta')$

とすることができる．

周期関数 f(x) を離散化し，ベクトル{v}を得，Ω区間で定義されるパターン{P}の集合によって，

$f(x) \fallingdotseq \{v\} \subseteq \{P\}$

とすることができる．

［条件A］{P}の各要素が相互に異なるとき，

［条件B］P_iと任意の実数値ξを一対一対応にする関数Ξが存在するならば，

$\xi_1 = v_{i+1}, \quad \ldots, \quad \xi_i = v_n$

とすることにより，

$\{P_1, \ldots, P_{n-1}\}$からP_nを計算することができる．P_nが$\{P_{n-1}\}$のあるパターンP_jに同じならば(この条件は条件Aと{周期＋Ω}分以上のデータがあれば満足される)

$P_{n+1} = \{P_n; \quad -v_{n-i+1} \quad +v_j\}$

である．ここで{}の中の;の右側の-/+記号は集合の要素を除く/追加することを表す．同様にして$\{P_{n+2}, P_{n+3}, \ldots\}$を計算できる．

これを関数の巡回表現と呼ぶ．巡回表現には見掛け上関数の「周期」に関する情報が入っていない．

3. 階層型ネットワークの特徴

階層型ニューラルネットワークは3層構造が本質的な構造である．この構造と情報の伝達機構を説明する．要素n個の入力ベクトルxがあるとき，第1,2層間において，

$x \equiv \{x_1, x_2, \ldots x_n\}$

$y_j = \sum V_{ji} x_i$,

$p_j = f(y_j)$

なる演算を考える．ここでfは一回微分可能な関数である．この関数を(ニューロンの)動作関数という．行列V_{ji}を(ニューロン間の)結合荷重，あるいは結合行列という．この結合荷重を決定する処理を(ニューラルネットワークの)学習という．第2,3層間における演算は， $p \equiv \{p_1, p_2, \ldots, p_m\}$

$q_k = \sum W_{kj} p_j$,

$o_k = g(q_k)$

$o \equiv \{o_1, o_2, \ldots o_l\}$

である．行列W_{kj}は第2,3層間の結合荷重，oがニューラルネットの出力である．

結合荷重V_{ji}, W_{kj}を決定するものを学習方程式という．学習方程式の中に誤差逆伝搬法がある．[3]

以上は1個のデータについての情報伝搬を示したが，階層型ニューラルネットの特徴は同様の情報伝搬が複数可能なことである．即ち，

$x \rightarrow \{x_\mu\}$, $p \rightarrow \{p_\mu\}$, $o \rightarrow \{o_\mu\}$

のようにベクトルをベクトルの集合に置換し，

その各々の要素であるベクトルについて上記の情報伝搬が可能である．階層型ニューラルネットワークの本質は，

$\{x_\mu\} \Leftrightarrow \{o_\mu\}$

なるベクトル集合間の対応をつける関数である．この対応において，対応は一体一対応ではではない．

$\{o_\mu\} = \{o_\nu\} = \ldots\ldots = \{o_\xi\}$

が許容される．しかし，

$\{x_\mu\} \neq \{x_\nu\} \neq \{x_\xi\}$

である．

3層の階層型ニューラルネットワークの偏微分形は，

$\partial o_k / \partial x_i = \Sigma W_{kj} g'(q_j) V_{ji} f'(y_j)$

である．[4] ベクトル表示では，

$\partial o / \partial x = W g' V f'$

である．

動作関数f, gはシミュレートする現象により選択する必要がある．一般的なsigmoid関数は，測定値が急激に変化しない周期的な現象を表現するのに適している．

4. 予測精度

第2節の予測理論を時系列現象に適用する，とき，一般には{周期＋Ω}分以上のデータは無い．（条件Ａは満足している）

従ってP_nが$\{P_{n-1}\}$のあるパターンP_jと同じにはならず，$\{P_1, \ldots, P_{n-1}\}$の各パターン間で距離：

$d_{nj} = \{\Sigma (P_n - P_j)^2\}^{1/2} \neq 0$

が計算でき，$\min\{d_{nj}\}$も必ず存在する．

今，$\min\{d_{nj}\} = d_{n\nu}$と表記する．

$$P_n - P_\nu = \{v_{n-i} - v_{\nu-i}, \ldots, v_n - v_\nu\} \equiv \{\delta_{n-i}, \ldots, \delta_n\}$$

であるから，

$$\begin{aligned}\{o_n\} &= \{o_\nu\} \\ &+ (1!)^{-1}(\partial o / \partial x)\{\delta_{n-i}, \ldots, \delta_n\} \\ &+ (2!)^{-1}(\partial^2 o / \partial x \partial x) \cdot [\{\delta_{n-i}, \ldots, \delta_n\} \times \{\delta_{n-i}, \ldots, \delta_n\}] \\ &+ \ldots \\ &\equiv \{o_\nu\} + \{\Delta\}\end{aligned}$$

である．ここで"×"は外積を表す．予測の信頼性は上記ベクトルの内積$\|\Delta\|(>0)$に比例すると考えられる．それは予測精度に対応する量である．

5. 常識の導入

時系列データが少ないとき，情報量の少なさを「常識」で補おうとするであろう．従来は予測にそういう恣意的な処理をすること自体が強く否定されて来た．それは予測値が変化してしまうからである．しかし，どのように変化するのかということ自体は検討に値する．

常識の例を示す．$P_n \equiv \{v_{n-i}, \ldots, v_{n-1}, v_n\}$なるパターンが$\xi_i = v_n$と対応しているとき，$P_\zeta \equiv \{v_{n-i} + \varepsilon, \ldots, v_{n-1} + \varepsilon, v_n + \varepsilon\}$と$\xi_\zeta = v_n + \varepsilon$を対応させる．そのような対応は一般的には成立しない．しかし，そういう対応を考えることは特に奇異なことではなく，我々は良くそういう対応を考えていることに気づく．

それは，常識から新しいデータを生成することである．（捏造する，と言うべきか？）

このとき，常識導入の是非はどう判定すべきか．データζを追加したとき第4節の予測精度が$\|\Delta\| \rightarrow \|\Delta_\zeta\|$と変化したとする．

このとき$\|\Delta_\zeta\| < \|\Delta\|$ならば常識導入の効果はあったと考えることが多いが，より多くのデータを使って，より正確な予測をしたわけであるから，その考えは正しくない．

$\|\Delta_\zeta\| > \|\Delta\|$となったときは，明らかにその常識導入は棄却すべきものである．第4節の予測にはそういう利用法がある．

別の常識を考察する．ニューラルネットワークの入力データ集合$\{x_\mu\}$はベクトルの集合であるが，ベクトルの各要素の値は区間[0, 1]に存在する．ゆえに無限多値論理演算[5] {|, &, ~}をその各要素に作用させることが考えられる．そのような操作はどういう作用を予測に与えるか考察する．

$$x_\mu = \{v_1, v_2, \ldots, v_k\}$$
$$x_\nu = \{v_1', v_2', \ldots, v_k'\}$$
$$x_\mu \mid x_\nu = \{\max(v_1, v_1'), \max(v_2, v_2'), \ldots, \max(v_k, v_k')\}$$

ここでｘと対応する教師データをξ_μ, ξ_νと

する．教師データについても同様な無限多値論理演算が可能である．従って，

$\xi_\mu \mid \xi_\nu = \max(\xi_\mu, \xi_\nu)$

このとき $v_1 < v_1'$ かつ $v_k > v_k'$ とする．添字1, kは時刻について言えば1<kである．教師データは時刻についてv_kと比較すると$v_k < \xi$である．そしてほとんどの場合$\xi = \xi_\mu$である．もしそうでないなら，時刻kの直後に突発的な現象が起ったことになる．以上をまとめると，2個の学習データを合成したデータは，予測値に関しては元のデータのどちらか一方である．それは第4節の結果と反する．

無限多値論理演算は{ | ,&,~}ばかりではなく，色々な演算が発表されている．たとえば確率的議論から{(+), ×,~}が定義されている．ここで(+)は代数的論理和，×は代数的論理積である．

$x_\mu (+) x_\nu = \{(v_1 + v_1' - v_1 * v_1'), (v_2 + v_2' - v_2 * v_2'), \ldots\ldots, (v_k + v_k' - v_k * v_k')\}$

こういう演算が是認されるのは$0 \leq v_i \leq 1$である必要がある．ニューラルネットで扱うデータはスケーリング処理により，その条件を満たしている．同様に

$\xi_\mu (+) \xi_\nu = \xi_\mu + \xi_\nu - \xi_\mu * \xi_\nu$

である．従ってξの小さい方の値も無視されない．それは第4節の偏微分係数($\partial o / \partial x$)が常に正であることに良く似る．

ただし代数的論理和は同一パータンに繰り返すと$x_\mu = \{v_1, v_2, \ldots, v_k\} \rightarrow 1$となる．(べき等律が成立しない．)教師データについても同様である．その性質は適当でない．無限多値論理で定義されている，その他の演算子(限界和，激烈和)について検討したが，べき等律において，適切でない性質を持っている．

従って常識導入による学習データを増加させる処理は非常に注意して行わなければならない．複数回の導入は厳に禁止すべきである．

普通，無限多値論理では否定演算子~を使う．そもそも論理を成立させるために，空間を定義し，論理値の否定が定義可能であるようにした．しかしニューラルネットワークの学習データ集合では，ある学習データの否定は存在しなくても良い．従って，学習データ集合間の論理的演算に否定演算子が現れないのは当然である．ゆえに否定演算子を含む論理演算で新たなデータを作り出すことはできない．逆に学習データの逆元が存在することが確実なとき，例えば現象が周期関数で表わされることが明らかなときは，その性質を否定演算子を常識として導入することにより，不足する学習データを補いうる．

6. まとめ

(1)本論文で示した階層型ニューラルネットワークと関数の巡回表現を用いた時系列現象の予測法は，関数形を部分断片に分解し，断片を組み合わせて外挿時の関数の形を予測するという原理に基づいている．従って断片の中に関数の将来の形と同じものが存在すれば精度良く予測可能である．

(2)ニューラルネットワークの出力値を入力値で偏微分した式から予測精度に対応する量を示した．

(3)ニューラルネットの学習データが不足して予測精度が不足するとき，データの不足を常識で補い予測精度を向上させる際の問題点を考察した．

参考文献

1)I. YOSHIHARA, M. NUMATA, T. AOYAMA, T. HARADA, "Prediction of Time Histories of Seismic Ground Motions using Genetic Programming", '99 KACC (1999.10.14).

2)青山智夫，井須芳美，長嶋雲兵「階層型ニューラルネットワークによる時系列現象の予想」，情報処理学会研究報告, 95-HPC-59, pp.13 (1995.12.11).

3)Rumelhart, D.E., McClelland, N.L., Eds., "Parallel Distributed Processing Exploration in Microstructure of Cognition", Cambridge, MA, 1986, Vols.1, Chapter 8.

4)青山智夫，市川紘「条件つきバックプロパゲーション学習法」情報処理学会研究報告, 91-NA-37, pp.59(1991.7.17).

5)向殿政男「ファジイ論理」日刊工業新聞社 (1993.3).