

ニューラルネットワークの複雑時系列予測への応用

Application of Neural Networks to Complex Time Series Prediction

宮野尚哉/Takaya Miyano・未来技術研究所 エレクトロニクス基盤研究部 デバイス技術室 主任研究員

紫富田 浩/Hiroshi Shibuta・和歌山製鉄所 製鉄部 製鉄工場 工場長

中嶋 研/Ken Nakashima・システムエンジニアリング事業部 和歌山システム部

池永泰治/Yasuharu Ikenaga・システムエンジニアリング事業部 ソフトウェアプロダクト部 営業・マーケティング室 室長

要 約

高炉で観測された時系列データには決定論的カオスと考えられるものが存在することを、最新の時系列解析手法を用いて明らかにした。この解析結果を先見知識として利用すると、時系列予測モデルとしてのニューラルネットワークの構造最適化が容易になる。不規則、乱雑に変動する時系列の近未来における挙動はニューラルネットワークを用いて予測できる。これらの成果は複雑に変動するシステムやプラントの制御、管理のための一つの新しい方向を示唆していると言える。

Synopsis

The possible existence of deterministic chaos in a blast furnace is investigated with nonlinear time series analysis for observed sequences of thermal fluctuations of the system. The analysis is exploited to acquire prior knowledge for determining the optimal architecture of neural networks as predictors of the observational time series. The networks are capable of learning the underlying dynamics and making short-term predictions about the series. The present work may show a novel approach to controlling complex systems.

1. 序 論

離散的な時間に対応する数値列を時系列という。観測により得られるデータは、測定に固有の時間分解能があるから時系列である。時系列解析はシステムやプラントの内部状態を診断するのに用いられる標準的な手法の一つである。信号の重要な特徴が規則的な変動にあるならば、データの特徴はパワースペクトル上で鋭いピークとして検出され、信号のエネルギー密度はピーク高さで計ることができる。時系列の挙動の将来予測も比較的容易である。しかしながら、時系列が不規則に変動する場合には、スペクトル解析は有効に機能するとは限らないし、時系列の将来傾向を予測することも容易ではない。

従来、不規則な変動は、自由度が非常に大きいシステムにおいて少数の自由度に関する情報やデータしか利用できないことからくる情報の欠如によって生起すると考えられてきた。しかしながら、近年著しい発展を遂げた非線形システム理論によると、少数の自由度からなる決定論的システムにおいてさえ、不規則で長期予測不可能な変動が起こり得るのである。このようなシステムの挙動はカオスと呼ばれている。幸いなことに、カオス系の将来変動は、近未

来に限るならばある程度予測可能である。どの程度予測可能かは、変動の決定論的側面がどの程度認められるかによる。この事実を応用すると、複雑、不規則に変動するシステムをより良い精度で制御、管理する技術の開発への糸口を見い出すことができるかも知れない。

本報告では、高炉で観測された不規則な時系列データを事例として、時系列の複雑さを解析し、解析結果を先見知識として用いて時系列の短期予測モデルを構築する手法について報告する^{1),2)}。高炉で観測された時系列には、少数自由度の決定論的カオスに相当すると考えられるデータが存在することが示されるであろう。最近数年のうちに開発された時系列解析手法を用いて、時系列の自由度に関連する埋め込み次元と、有効な予測時間間隔を推定する³⁾。これらの先見知識のもとで、高炉データの短期予測モデルを構築する。予測モデルのベースとなるのは、一般化動径基底関数ネットワークと呼ばれる広義のニューラルネットワークである。このシステムは、時々刻々、計測、収集されるデータのオンライン学習により随時最適化されるという利点がある。

2. 理 論

多自由度系システムの1自由度に関する時系列データだけを利用して元の系の時間発展のダイナミクスを解析することを可能にするのが、時系列の埋め込み(embedding)である^{4),5)}。N個のデータ点からなる時系列 $\{x(t)\}_{t=1}^N$ が与えられたとする。x(t), x(t-Δt), ..., x(t-(D-1)Δt)をD個の独立変数と考えて、D次元空間で時系列から軌道を再構成する。Dは埋め込み次元、Δtは適当な時差である。Dが適切に選択されれば、元のシステムの力学的性質は埋め込み空間で再構成された軌道を包む多様体にそっくり移植されるので、このような変数変換が意味であるものとなる。もし、時系列に決定論的側面が認められるならば、関数近似を利用して時系列の現在および過去の値から、未来の値を予測できる。

$$x(t+\tau\Delta t) = f[x(t)] + \varepsilon(\tau) \quad \dots\dots\dots(1)$$

$$x(t) = (x(t), x(t-\Delta t), \dots, x(t-(D-1)\Delta t)) \quad \dots\dots\dots(2)$$

ここで、τは予測時間ステップ、ε(τ)は予測誤差を表す確率変数である。本報告では、予測値と実測値との相関係数γ(τ)、および、予測値と実測値との間の平均自乗平方根誤差を元の時系列の標準偏差で規格化した値E(τ)で予測誤差を評価する。予測モデルを構築するにあたって最初に必要となる情報は、予測モデルの入力ベクトル次元である埋め込み次元Dと、予測が有効な時間間隔τである。次に重要な情報は、時系列変動のダイナミクスが線形相互作用か非線形相互作用かということである。非線形ダイナミクスによる時系列の予測モデルには、ニューラルネットワークのような非線形近似モデルを利用すべきであろう。不規則で長期予測不可能な変動には、非線形ダイナミクスから生じるカオスと、確率的線形ダイナミクスから生じる自己相似ランダムがある。自己相似ランダムはカオスに似た挙動を示すので少数自由度のカオスと混同されやすいが、時系列の時間スケールをκ倍すると同時に時系列強度のスケールをκ^{-H}倍しておくとき時系列の統計量が保存されるという特有の統計的性質を持っている⁹⁾。Hはスケーリング係数と呼ばれる定数である。スケール変換を適切に行なうと、どの時間スケールで眺めても統計的性質が同じに見えることが「自己相似」の名の由来である。自己相似ランダムは強い線形相関を示すので、その予測モデルには線形自己回帰が適切である。したがって、不規則な時系列がカオスか自己相似ランダムか同定することは、予測モデルの選択にかかわる重要な問題である。

本報告では、高炉データの予測モデルとして一般化動径基底関数ネットワークを用いるが、ネットワーク構造を決める先見知識を獲得するにはSugiharaとMayにより開発された時系列予測法を利用する⁷⁾。Sugihara-May法の概略を述べる。時系列を前半(L)、後半(P)に二分し、各々について入力ベクトル—出力値対(x(t), x(t+τΔt))を生成

する。Lに属する入出力対を時系列の挙動を代表する参照パターンと考える。x(t_p) ∈ Pに対する未来値を予測するために、D次元埋め込み空間内でx(t_p)を取り囲むD+1多面体であって、その頂点にx(t_i) ∈ Lがあるものの中から最小の多面体を見つけ、x(t_p)と各頂点のx(t_i)とのユークリッド距離d_{ip}を計算する。x(t_p)に対する将来値y(t_p+τΔt)は、

$$y(t_p + \tau\Delta t) = \frac{\sum_{i=1}^{D+1} x(t_i + \tau\Delta t) \exp(-d_{ip})}{\sum_{i=1}^{D+1} \exp(-d_{ip})} \quad \dots\dots\dots(3)$$

$$d_{ip} = |x(t_i) - x(t_p)| \quad \dots\dots\dots(4)$$

により予測される。現在の挙動パターンに類似した過去の挙動パターンから未来を予測するのである。この予測法はニューラルネットワークのように学習プロセスを必要としないので、データ量があまり多くないならば比較的短時間で予測を実行できる。最適なDを見つけるには、τ=1と固定して様々なDについて予測を行ない、予測誤差が最小となるDを求めればよい。また、最大有効予測時間を求めるには、最適なDに固定して今度は様々なτについて予測を行ない、予測誤差が許容限度内となるような最大のτを見つければよい。これらの予測手続きはそれほど手間がかからない。一方、時系列がカオスか自己相似ランダムか見分けるには、予測誤差の予測時間依存性を調べればよい。時系列がカオスならば、

$$\frac{E(\tau)}{E(1)} = \frac{\exp(\lambda\tau\Delta t)}{\exp(\lambda\Delta t)} \quad \dots\dots\dots(5)$$

が成立する⁹⁾。これは、初期値の僅かな差が指数関数的に増大するというカオスの特徴を表す。λは最大リアプノフ数であり、情報の損失率に相当する。時系列が自己相似ランダムならば、

$$\frac{E(\tau)}{E(1)} = \tau^H \quad \dots\dots\dots(6)$$

が成立する^{9),10)}。Hはスケーリング係数である。これらの関係式によると、log[E(τ)/E(1)]をτまたはlogτについて片対数および両対数プロットを取り、線形相関を示すのが片対数プロットならば時系列はカオス、両対数プロットならば自己相似ランダムということになる。こうして、時系列のダイナミクスが決定論的非線形か確率的線形か知ることができる。

一般化動径基底関数ネットワークは、動物の視覚野の機能¹¹⁾をモデル化するために、PoggioとGirosiによって開発された関数近似手法である¹²⁾⁻¹⁴⁾。本研究で用いるネットワークは、

$$f[x(t)] = \sum_{h=1}^M c_h \exp[-\beta_h |x(t) - \theta_h|^2] \quad \dots\dots\dots(7)$$

と表される。c_h, β_h, θ_hは学習によって最適化されるパラメータ、Mは基底関数として使用するガウス関数の個数である。神経生理学的には、ガウス関数は視覚野においてカラム構造を形成する神経細胞集合体の1ユニットに対応するであろう。各ユニットの集合が視覚野のネットワークを

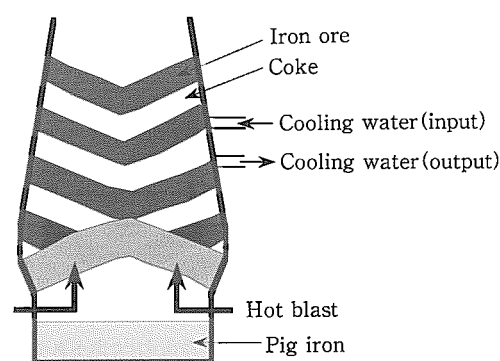
形成するのと同様に、ガウス関数の集合が動経基底関数ネットワークを形成する。ネットワークの学習則は以下のとおり。Sugihara-May 予測法の場合と同様に、時系列を前半、後半に二分し、各々について入力ベクトル—出力値対を作成する。前半時系列を学習パターンとして利用する。学習用入力ベクトルを(7)式に入力し、その出力値 $f[x(t)]$ が学習用出力値 $x(t+\tau/2)$ にできるだけ近くなるように、パラメータ c_h , β_h , θ_h を変化させる。パラメータの調節には確率的勾配降下法を用いた。本報告ではこの学習アルゴリズムの詳細は述べない。詳細は参考文献^{15),16)}を参照されたい。基底関数の個数 M は試行錯誤で決定する。ネットワークの学習に期待することは、最適化されたパラメータで決まる基底関数の組み合わせで時系列のダイナミクスを再現することである。つまり、入力ベクトルで表現される現在および過去の挙動と、出力値で表現される未来の挙動との間の一般的な関係を、基底関数の適当な組み合わせで再現するのである。学習用入出力対が十分な量利用できるならば、このようなことが可能であろう。学習が適切に行なわれたならば、学習時には現れなかった未知の入力ベクトルをネットワークに入力すると、その未来値が適切に予測される。パラドクシカルな表現をするならば、学習は、学習しなかったことに対して正しく対処するために行なうのである、ということになるだろう。(7)式は入力値に対する非線形応答特性を表現するので、カオス時系列のような非線形ダイナミクスを近似するには効率がよい。しかしながら、確率線形過程の近似には効率が悪い。カオス時系列に含まれている観測ノイズや自己相似ランダム時系列に相当する時間変動を効率よく近似するには、線形自己回帰が適切である。本報告では、線形自己回帰モデルを並列に付加した動経基底関数ネットワークを用いる^{11),16)}。

$$f[x(t)] = \sum_{h=1}^M c_h \exp[-\beta_h |x(t) - \theta_h|^2] + a \cdot x(t) + b \quad \dots\dots\dots (8)$$

a , b は線形自己回帰パラメータである。(8)式の最適化方法は以下のとおり。時系列前半を学習パターンに利用し、最小自乗近似で線形自己回帰パラメータを決める。次に、線形自己回帰モデルの残差とネットワーク出力値との誤差ができるだけ小さくなるように確率的勾配降下法により c_h , β_h , θ_h を最適化する。この予測モデルでは、データに含まれている非線形成分(カオス成分)は動経基底関数ネットワークで近似され、外部ノイズ等の線形成分は線形自己回帰モデルで近似される。観測データは必ずある量の外部ノイズで汚染されているから、このネットワークアーキテクチャは実際のデータを扱うのに適していると言えるだろう。動経基底関数ネットワークは、Sugihara-May 法に比べ、学習に多大の時間を要するが、一旦学習が終了すると非常に高速に予測を実行できるという長所がある。

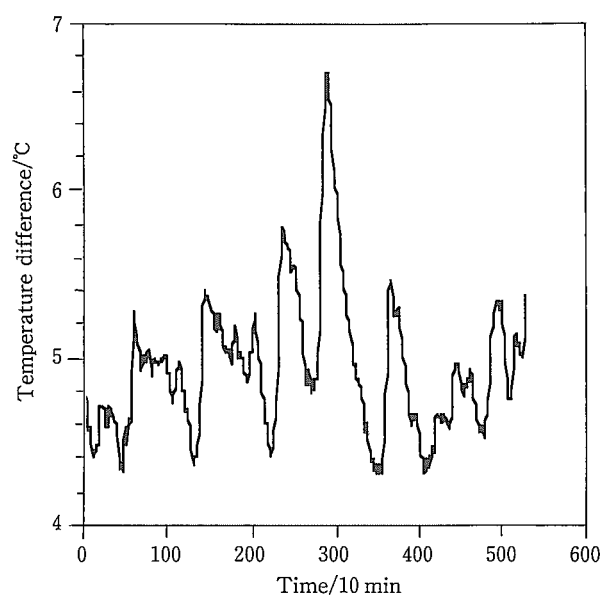
3. 高炉データの時系列解析

高炉側壁には、高炉鉄皮を熱ダメージから保護するためにステープ冷却盤と呼ばれる水冷冷却盤が多数取り付けられている。各ステープ冷却盤の冷却水給排水温度差は高炉操業における重要なデータである。第1図に高炉断面の概念図を示す¹⁷⁾。第2図は高炉操業介入のない期間に計測された10分ごとのステープ冷却水給排水温度差変動(データ点数=528)である¹⁾。操業介入のない期間における温度差データを短時系列、操業介入を多数含む期間における温度差データを長時系列(データ点数=8189)と呼ぶことにする。



第1図 高炉断面の概念図¹⁷⁾

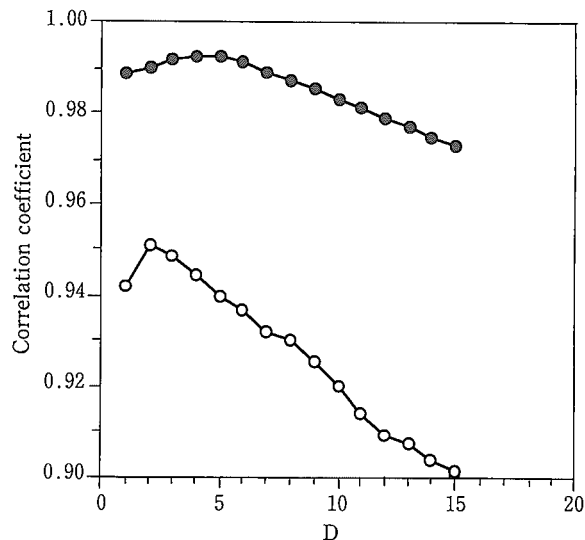
Fig.1 Schematic cross section of a blast furnace¹⁷⁾



第2図 高炉操業介入のない期間におけるステープ冷却水給排水温度差変動¹⁾

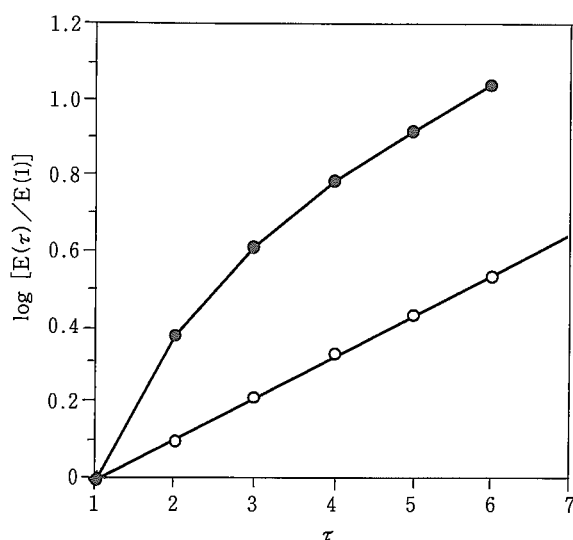
Fig.2 Short time series of temperature differences between input and output cooling water of a cooling apparatus¹⁾

いずれの時系列も、パワースペクトルにはピークが現れないので、不規則な変動を表す。これらの時系列について、時系列前半を参照パターンとして使用し、時系列後半を Sugihara-May 法により予測した。第 3 図は $\tau=1$ と固定した場合の予測における予測誤差 $\gamma(1)$ の埋め込み次元依存性を表す。



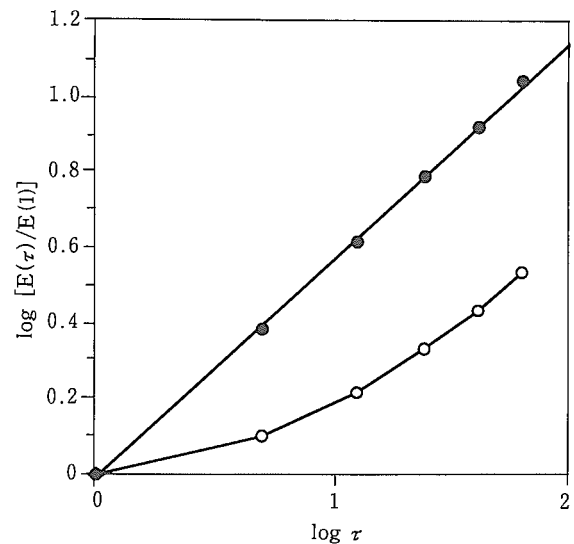
第 3 図 Sugihara-May 法による予測における予測誤差 $\gamma(1)$ の埋め込み次元依存性 ($\tau=1$, $\Delta t=10 \text{ min}$)
●: 操業介入あり, ○: 操業介入なし

Fig.3 Correlation coefficient between predicted and actual values as a function of embedding dimension for the long time series (●) and the short time series (○)



第 4 図 $\log[E(\tau)/E(1)]$ - τ プロット ($D=3$, $\Delta t=10 \text{ min}$)
●: 操業介入あり, ○: 操業介入なし

Fig.4 Plot of $\log[E(\tau)/E(1)]$ versus τ for the long time series (●) and the short time series (○)



第 5 図 $\log[E(\tau)/E(1)]$ - $\log \tau$ プロット ($\tau=1$, $\Delta t=10 \text{ min}$)
●: 操業介入あり, ○: 操業介入なし

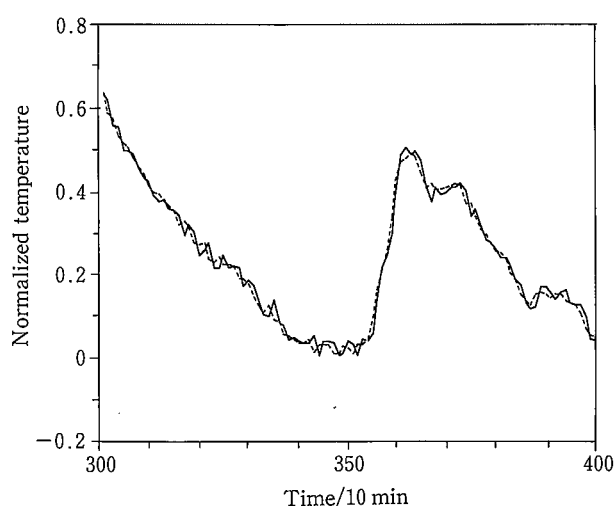
Fig.5 Plot of $\log[E(\tau)/E(1)]$ versus $\log \tau$ for the long time series (●) and the short time series (○)

短時系列, 長時系列いずれも $D=3$ 付近が最良予測に対応すると考えられる。この結果から, 予測モデルの埋め込み次元は $D=3$ と決めた。 $D=3$ における予測精度の τ 依存性からは, 実用的に予測が有効と見做せる予測時間間隔は $\tau \leq 3$ であると推定された。

第 4 図および第 5 図はそれぞれ, $D=3$ における $\log[E(\tau)/E(1)]$ - τ プロットおよび $\log[E(\tau)/E(1)]$ - $\log \tau$ プロットである。短時系列は片対数プロットが良い線形相関を示す。相関の線形性の程度を示す相関係数は 0.999 である。片対数プロットの傾きから推定されるリアプノフ指数は $\lambda=0.108 [0.1 \text{ min}^{-1}]$ である。一方, 多数の操業介入を含む長時系列は両対数プロットが良い線形相関を示す。線形相関の程度を評価する相関係数は 0.999, 両対数プロットの傾きから推定されるスケーリング係数は $H=0.577$ である。これらの結果は, ステープ冷却水温度差時系列は操業介入のない期間では決定論的カオスに見えるが, 操業介入が加わることにより線形確率過程から生じる自己相似ランダム時系列に見えることを示唆している。時系列の本来のダイナミクスは決定論的非線形でカオス的であるが, 操業介入が外部ノイズとして頻繁に作用することによりダイナミクスの非線形性が隠されたと解釈できる。予測モデルとしては次節で述べる線形自己回帰モデルを並列に付加した動経基底関数ネットワークが適切であろう。動経基底関数項は時系列本来の非線形ダイナミクスを再現すると期待され, 線形項は操業介入による線形確率過程を近似するのに効果があると考えられるからである。このような予測モデルが有効に機能することを次節で実証する。

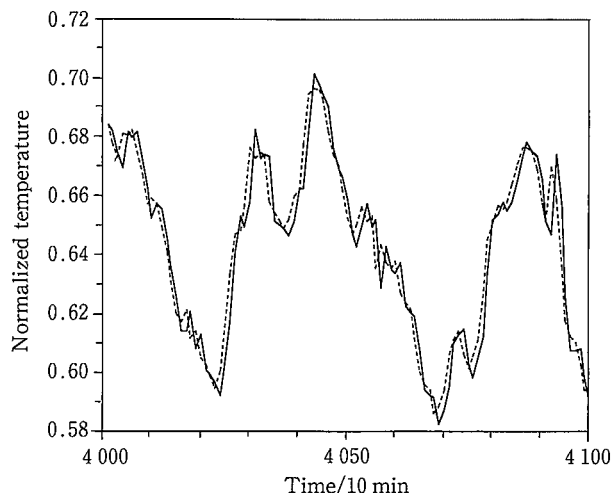
4. 高炉データの時系列予測

$D=3$, $\tau=1$ のもとで、線形自己回帰モデルを並列に付加した動径基底関数ネットワークにより高炉データ予測モデルを構築する。短時系列の前半 300 点を学習用データに用いてネットワークの学習を行なった。試行錯誤法により、ガウス関数の個数、すなわち、ネットワークの中間層ノード数 $M=2$ と決めた。学習後の動径基底関数パラメータ c_h と線形自己回帰パラメータ a_i の絶対値はほぼ同じ大きさであった。したがって、動径基底関数項および線形自己回帰項のいずれからも同程度の寄与があると言える。第 6 図はこの予測モデルによる予測結果である。図の縦軸の温度スケールは最大値と最小値の差で規格化されている。予測誤差は $E(1)=0.134$ であり、予測値は実測値とよく一致していると言える。一方、長時系列については、 $D=3$, $\tau=1$ は同じであるが、時系列前半 4000 点を学習データに用い、中間層ノード数を増やし $M=3$ とした。ネットワークパラメータ数 w 、学習パターン数 n 、学習誤差 ε の間には、 $n=O(w/\varepsilon)$ なる関係が成立することが知られている¹⁸⁾。この関係が時系列予測においても成り立つと仮定すると、長時系列は短時系列よりも 10 倍以上多くの学習パターンを生成できるので、ネットワークパラメータを増やすことができる。これが長時系列において中間層ノード数を増加することができた理由である。第 7 図は長時系列に関する予測結果である。予測誤差は $E(1)=0.079$ であり、予測値は実測値とよく一致している。この場合にも、学習後の動径基底関数パラメータ c_h と線形自己回帰パラメータ a_i の絶対値はほぼ同じ大きさであり、動径基底関数項および線形自己回



第 6 図 線形自己回帰モデルを並列に付加した動径基底関数ネットワークによる短時系列予測結果 ($D=3$, $\tau=1$, $\Delta t=10 \text{ min}$, $M=2$, 実線: 予測値, 破線: 実測値)

Fig.6 Forecasts about the short time series with a generalized radial basis function network added in parallel to a linear autoregressive predictor



第 7 図 線形自己回帰モデルを並列に付加した動径基底関数ネットワークによる長時系列予測結果 ($D=3$, $\tau=1$, $\Delta t=10 \text{ min}$, $M=3$, 実線: 予測値, 破線: 実測値)

Fig.7 Forecasts about the long time series with a generalized radial basis function network added in parallel to a linear autoregressive predictor

帰項のいずれからも同程度の寄与があった。前節の時系列解析においては、短時系列と長時系列では統計的性質が異なっていたが、いずれの時系列にも非線形ダイナミクスによる変動成分が含まれていると考えてよいであろう。

5. 結 論

時系列予測モデルを構築する過程で、データのダイナミクスを解析することにより、埋め込み次元や最大有効予測時間、データの非線形性を解析しておくことは予測モデルの構造を決める先見知識を得るという観点から重要である。本報告では Sugihara-May 法に基づいて時系列解析を行なった。この手法は、利用可能なデータ量が数百点程度の少数であっても比較的信頼性の高い検定結果を導くので有用である。計測によって得られた時系列データは、測定ノイズやシステムへの制御介入に起因したあるレベルの外部ノイズに汚染されている。このような時系列データを扱うには、本報告で紹介したような線形自己回帰モデルと動径基底関数ネットワークの並列ハイブリッド型予測モデルが有効である。線形項は外部ノイズのダイナミクスを再現し、非線形項はシステム本来の非線形ダイナミクスを再現する。外部ノイズレベルが減少したり、システムの非線形ダイナミクスが消失した場合には、オンライン学習時に線形パラメータまたは非線形パラメータが自動的に 0 になる。この意味で本報告におけるネットワークアーキテクチャは実際的であると言えるであろう。



宮野尚哉 / Takaya Miyano

未来技術研究所 エレクトロニクス基盤研究部
デバイス技術室 主任研究員

(問合せ先: 06(489)5743)

参考文献

- 1) 宮野尚哉, 紫富田 浩, 中嶋 研, 池永泰治: 電子情報通信学会論文誌, **J 79-A**(1995), p.38
- 2) T. Miyano and K. Aihara: *Proc. Int. Symp. on Artificial Life and Robotics AROB'97*(1997), p.150
- 3) 宮野尚哉, 数理科学: 「カオス研究の最前線 15: カオスと時系列解析」, 発表予定
- 4) F. Takens: *Lect. Note in Mathematics* **898**(1981), Springer-Verlag
- 5) T. Sauer, J. A. Yorke, and M. Casdagli: *J. Stat. Phys.*, **65**(1991), p.579
- 6) A. R. Osborne and A. Provenzale: *Physica D*, **35**(1989), p.357
- 7) G. Sugihara and R. May: *Nature*, **344**(1990), p.734
- 8) M. Casdagli: *Physica D*, **35**(1989), p.335
- 9) A. A. Tsonis and J. B. Elsner: *Nature*, **358**(1992), p.217
- 10) T. Miyano: *Int. J. Bifurcation and Chaos*, **6**(1996), p.2031
- 11) N. K. Logothetis, J. Pauls, and T. Poggio: *Current Biology*, **5**(1995), p.552
- 12) T. Poggio and F. Girosi: *Proc. IEEE*, **78**(1990), p.1481
- 13) T. Poggio: *Cold Spring Harbor Symposia on Quantitative Biology*, **LV**(1990), p.899
- 14) F. Girosi, M. Jones, and T. Poggio: *Neural Computation*, **7**(1995), p.219
- 15) T. Miyano, H. Morita, A. Shintani, T. Kanda, and M. Hourai: *J. Appl. Phys.*, **76**(1994), p.2681
- 16) T. Miyano and F. Girosi: *A.I. Memo*, **No.1447**(1994), Massachusetts Institute of Technology, Artificial Intelligence Laboratory
- 17) 合原一幸: 「カオスの数理と技術」(1997), p.91-93, 放送大学教育振興協会
- 18) E. B. Baum and D. Haussler: *Neural Computation*, **1**(1989), p.151