

カオス時系列予測における複数のベクトル次元を用いた 予測精度向上に関する試み

Attempt Concerning Forecast Accuracy Improvement
that uses Two or More Vector Dimensions in Chaos Time Series Forecast.

渡邊 達男
Tatuo WATANABE

1. 序論

自然界のさまざまな現象は複雑で、その多くは予想するのが大変困難である。もし台風や、地震、その他の自然現象を予測できれば、災害や、それに伴う人的被害も回避することが可能性があり、切望されるところである。しかし実際には自然界はとても複雑で、特に長期の予測は困難を極める。

最近の非線形科学は少しずつであるが、困難な自然界の予測を行おうと試みている^{1,2,3)}。様々な方法で予測を行おうとしているが、複雑な自然界を全体的に予測するのは難しい。しかし、自然現象をモデル化し、モデルを簡単にし、1変数の関数として、さらには短期の予測するのはそれほど難しくないと最近知られるようになってきた。

自然界の予想では、古くはLorenzが気象現象を予測する為に、3変数の予測関数系、いわゆるLorenz系を提案している³⁾。この系は非線形の3変数の微分方程式であり、残念ながら気象現象を予測するのには不足していたが、後のカオス系の研究の火種になり、多くの提案をした。

時系列予測に関してはその後、幾つかの予測器が考えられている^{1,2)}。その幾つかは以前の筆者の論文で考察した^{4,5,6,7)}。その一つとしてLorenzの類推法¹⁾では、数ステップ先までは予測できるが長期予測は非常に困難であると思われた。Lorenzの類推法を改良した方法が幾つか知られており、筆者も近傍点数、埋め込み次元数、等のパラメータを変化させることにより、予測精度が変化することが示された⁴⁾。

最近よく話題になっている動径基底関数ネッ

トワーク法^{1,2)}(RBFN法,Radial Based Function Network)では、大変少ない過去のデータで、比較的良い短期予測が可能であったが、複雑な自然現象、例えば台風の上陸数等は予測が難しかった。

この論文では、いままで筆者の論文で扱ってきた予測器を用いて、さらなる予測精度の改良が出来ないかを試みる。

それにはまず、Lorenzの類推法を用いて、いくつかのパラメータ値による予測値を平均化する等、同じ予測器でパラメータを変えた時の予測値を操作し、予測精度を向上することを試みる。

またRBFN法でも同様にパラメータを変えた予測値を元に、新たな予測値を作り出し、予測精度の向上を試みる。

2章は、Lorenzの類推法とRBFN法の解説に当てられる。3章はLorenzの類推法のパラメータを変えたときの予測値を用いて、予測精度が向上することを報告する。4章はRBFN法を用いて同様に予測精度が向上することを報告する。5章は考察に、6章はまとめに当てられる。

2. 予測器について

2-1 Lorenzの類推法

E.N.Lorenzは、1969年、類推法 (the method of analogues) と呼ばれる手法を提案した。

今、1次元時系列 $x(0), x(1), \dots, x(t)$ を要素に持った m 次元ベクトル $\mathbf{v}(t) = \{x(t-m), x(t-m+1), \dots, x(t)\}$ に対して、そのベクトル空間上で、今わかっている最後の点のベクトルの最近傍点の次の時系列ベクトルをもって、予測点 (まだ知ら

れていない次のベクトル)とした。これは今の時系列パターンと同じパターンを過去に見出し、過去のパターンが繰り返されるという仮定をもとに、その情報から予測を行うものである。

データにノイズが無い場合には非常に有効な予測手段であるが、ノイズの量が多いと、近傍点が果たして最も良い点であるかが問題となり、予測精度が悪くなる事が考えられる。しかし、最近傍点数を増やす事で、ノイズを低減させ、予測精度が向上する。複数の近傍点からどのように予測ベクトルを一つ決定するかにより、予測精度も変化する。もっとも一般的なのは複数のベクトルの加重平均を取ることである。すなわち、近傍点ベクトル $\mathbf{v}(k_i)$ 、近傍点数 N 、近傍点と最後の点のベクトルとの距離 d_i とすると、予測点ベクトル $\mathbf{v}(T)$ は、

$$\mathbf{v}(T) = \frac{\sum_{i=1}^N d_i \mathbf{v}(k_i)}{\sum_{i=1}^M d_i} \quad (1)$$

で与えられる。

この論文では、2個の近傍点を取り($N=2$)、(1)式を用いて、予測点のベクトルを決定する方法を選択した。

2-2 RBFN法

動径基底関数ネットワーク法(RBFN法)は、変形された2層ニューラルネットワークの一種で、動径基底関数という局在化された基底関数を重ね合わせることで、任意関数補完の方法として考えられて来た。以下に簡単に説明をする。

もともと関数補完の方法であったが、M.Casdagli⁸⁾はRBFN法をカオス時系列予測問題に応用した。

今、時系列 $x(0), x(1), \dots, x(t)$ を要素を持ったベクトル $\mathbf{v}(t) = \{x(0), x(1), \dots, x(t)\}$ に対して、次の式を考える。

$$x(t+1) = \sum_i^M \lambda_i \Phi(\|\mathbf{v}(t) - \mathbf{c}_i\|) \quad (2)$$

ここで、 λ_i は重み、 $\mathbf{c}_i$ はセンターと呼ばれる点、 Φ は基底関数である。

基底関数 Φ には幾つかのものが使われる。よく使われるものに、ガウシアン関数があり、

$$\Phi(r) = \exp(-r^2/b) \quad (3)$$

で与えられる。

シンプレートスプライン関数、

$$\Phi(r) = r^{2k+1} \log r \quad (4)$$

ベル型関数、

$$\Phi(r) = \frac{1}{1 + \cosh(br)} \quad (5)$$

等もある。今回はガウシアン関数(3)を用いて計算した。

RBFN法を用いて時系列を予測する場合には、あらかじめ既に知られている時系列(データと呼ぶ)を用いて重み λ_i を計算する。 λ_i は2層ニューラルネットの結合係数に相当するが、学習を用いて決定するのではなくて、一般には式(2)に知られているデータ時系列を代入し、連立方程式を解くと λ_i が求められる。

今回の報告では、過去のデータ時系列を用いて、 λ_i を決定する方法を用いた。

センター $\mathbf{c}_i$ は任意である。センターはベクトルで、一般的には時系列 $\mathbf{v}(t)$ と同じ次元でなくてはならない。センター $\mathbf{c}_i$ に関しては筆者の論文⁶⁾において、過去の時系列データを用いることで良い予測ができることがわかっている。今回は過去の時系列データをセンターに採用することにする。

上記の方法はスタンダードRBFN法と呼ばれているが、式(2)に線形項を加えたアフラインプラスRBFN(APBFN)もある。この方法はノイズ成分を線形項に吸収できるので、ノイズを含む時系列に有効である¹⁾。

$$x(t+1) = \mathbf{A}\mathbf{v}(t) + \mathbf{B} + \sum_{i=1}^M \lambda_i \Phi(\|\mathbf{v}(t) - \mathbf{c}_i\|) \quad (6)$$

3. Lorenzの類推法を用いた、複数パラメータの予測結果による予測

この章では、Lorenzの類推法を用いて、実際に予測を行ってみた。その際、単一のパラメータの組での予測、例えば、ベクトル次元 m や近傍点数 N を固定して予測した場合と、ベクトル次元を変化させて予測した予測結果を平均した予測結果とを比較し、複数パラメータによる予測結果の平均がどのように予測精度を向上させるかを数値実験により行った。以下にその結果を示す。

3-1 Logistic写像の予測

Logistic写像は以下のような式で表され、当初は生物の個体数の変動を表す式として提案された。

$$x_{n+1} = ax_n(1-x_n) \quad (7)$$

ここで、 a はパラメータで、 a の値により、収束、振動、カオスになる。 $a > 3.57$ ではカオスに成る可能性が高い。

ここでは、 $a=3.8$ 、 $x_0=0.5$ として計算を行い、 x_{101} から、それ以前のデータを用いて予測を開始した。その際、ベクトルの次元 $m=2$ から5までの予測値とそれらの予測値の平均も求めた。また近傍点数 $N=2$ とした。結果を図1に示す。

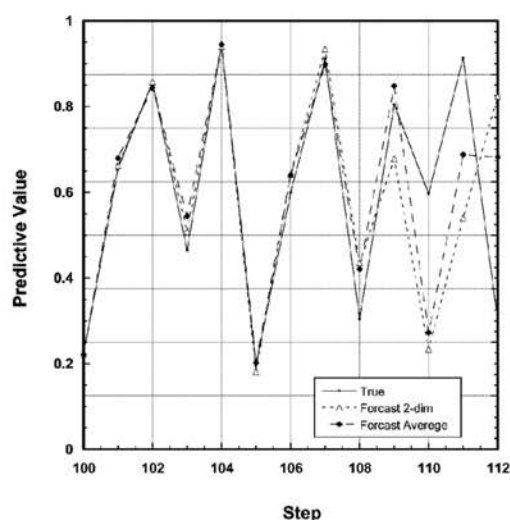


図1 Logistic写像のLorenzの類推法による予測結果。

図では、真値（(7)式を用いて計算した値）、 $m=2$ での予測値、 $m=2\sim 5$ までの予測値の平均予測値を示している。予測は101ステップから行っている。従って100ステップ目はすべて同一の値である。

図を見ると、107ステップまではほぼ予測がうまく行っているが、108ステップから予測がずれ始めている。しかし $m=2$ での予測と平均値での予測では、平均値での予測のほうが、やや真値とのずれが少ない。101ステップから112ステップまでの真値との相関係数を計算すると、

表1 予測値の相関係数

m	相関係数
2	0.67
2~5の平均	0.81

表1のようになり、予測の向上が見られることがわかる。

3-2 日本本土への台風上陸数の予測

毎年日本は台風による災害が発生する。

ここでは、日本本土に上陸する台風の数にLorenzの類推法で予測を試みた。元となるデータは1951年から1987年までを用い⁹⁾、1988年から1992年までの台風の上陸数を予測した。その場合、ベクトル次元を $m=2\sim 5$ まで変化させ、 $m=2$ による予測値と、それらの予測値の平均を予測値としたものを計算した。

結果を図2に示す。図2を見ると、平均した予測値が次元 $m=2$ の予測値に対して、多少向上していることがわかる。例えば、1988年はほぼ予測しているが、1989年には $m=2$ の予測値は急激に少なくなっている。しかし、平均した予測値は1988年に比べて増加している。また、1990年には予測値はどちらも同じような値になり、予測は外れてきている。

この結果から、幾つかのベクトル次元による予測値の平均を取ると、多少なりとも予測値の増加減少の傾向の予測の向上になっていると思われる。ただし、やはり自然界の予測は難しく、台風の上陸数にも何らかの数学的規則があると思われるが、そのパターンを把握し予測できて

いるとは言い難い。

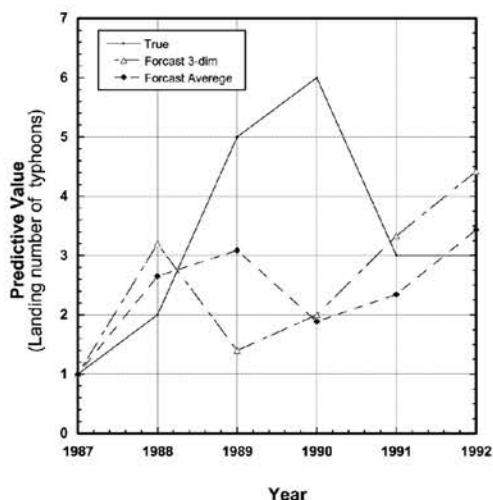


図2 Lorenz類推法を用いた、日本本土への台風上陸数の予測

4. RBFN法を用いた予測

同様に、RBFN法を用いて日本への台風の上陸数の予測を行った。RBFN法では、センターというベクトルが重要であるが、センターは元となる台風の上陸数の時系列そのものを用いた。そのうえで、時系列ベクトルの次元 m を変化させて予測し、予測値の平均を取り、単独の次元での予測値と比較してみた。

図3はセンター数 $M=20$ 、1951～1971年までのデータをセンターとして、ベクトルの次元 m を6～10まで変化させ予測した値の平均の予測値を示す。また $m=6$ の予測値も示してある。

図を見ると、1992～1994年は $m=6$ と平均のグラフとも予測がよく当たっている。しかし、1995年以降では、 $m=6$ のグラフは外れている。それに比較して、平均のグラフは値としてはずれているけれども、増減のパターンはほぼ予想が当たっているように見える。1990～1992年に関してはどちらが当たっているとも言えないが、1990年では真値と平均した予測値との差が、 $m=6$ の場合よりも少ない。従って、10年間全体としては平均した予測値の方が当たっているのではないと思われる。

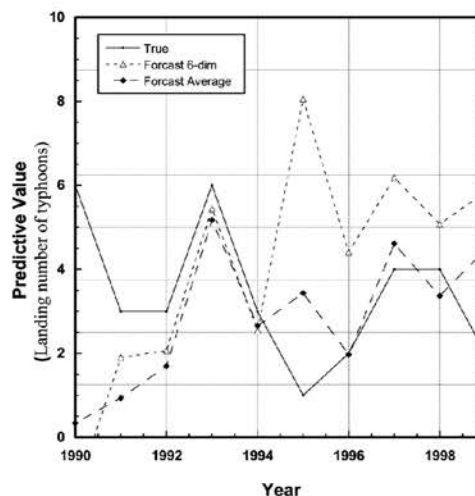


図3 RBFN法を用いた日本の台風の上陸数の予測

相関係数を計算すると、1990,1991,1999の3年を抜いた予測値で、表のようになる。

表2 台風の予測の相関係数

m	相関係数
6	-0.13
6~10の平均	0.65

5. 考察

時系列予測における、次元を変化させた時の予測値の平均値がどのくらい予測を向上させるかを調べてきた。その結果、予測結果を平均すると、予測が向上する場合があることがわかった。

しかし、実際の自然界の予測ではなかなか難しい所がある。例えば、Lorenzの類推法では、近未来は予測できるが、すぐに予測が悪くなる。しかし単独のパラメータだけよりは少し良くなる。もともとの単独のパラメータによる予測が誤差を含み、その誤差が真値から土の方向に均等にずれていれば、即ち例えば正規分布するような誤差が存在すれば、この平均化する方法により、予測値は向上する事が考えられる。しかし、どうもそう簡単ではないようである。即ち、それぞれの次元の異なった予測値では、パターンの読み違いによるずれであるので、平均化することが必ずしも予測値の向上を意味しない。パタ

一の読み違いによっては、逆に予想値が悪くなる可能性も考えられなくはない。この点についてはもっと多数の予測を実際に行ってみて、予想すべき時系列によってどのようになるかを見て行き、さらに、どのようにしてパターンの読み違いが統計的に修正可能か、数値実験、及び理論的に詳細に議論する事が必要であろう。

また、単純に予測値を平均するだけではなく、予測値をパラメータに依存した重みを付けて、予測値を計算する方法もあるだろう。その場合、どのような重み付けをどのように決定するかも考える必要があるだろう。

今回は行わなかったが、次元以外の他のパラメータを変化させ、それぞれの予測値を平均させる様々なやり方が残されている。例えば、RBFN法ではセンター数を変化させてみる等である。特に問題なのは、(3)式のガウシアン関数の b の値である。実際の予測においてこの b を変化させると、ある程度の予測値の変化が見られ、予測値の変化のパターンは変わらないが、予測値の絶対値の変化がおこる。この b の値による予想値の依存性は大きく、詳しく調べる必要がある。今回の数値実験では、最も当たっていると思われる b の値を人為的に選択し、予測を行っているので、 b の最適値は何なのかに対する不正確さが残っている。

予測器には様々なものがあり、予測器が異なるものに対する予測値の平均値をもって予測値とする方法も考えられる。例えば、Lorenzの類推法の予測値とRBFN法の予測値の平均を予測値とする方法である。また重みを付けて平均をしてもよい。この場合、重みの付け方をどのようにするかは試行錯誤で、予測が良いものも見つけて、その後、実際の自然界の予測を行う事も一つの方法であると思われる。

予測値の平均を扱う場合、次元によっては平均された予測値より予測が良かった場合がある。例えばRBFN法で $m=9$ の場合は相関係数が0.67であり、次元による予測値を平均した予測値よりも良い値である。ただし、単純にどの次元が予測が良いかわからない時は、平均すると予測値は良くなったと言える。

6. 結論

Lorenzの類推法とRBFN法の複数のパラメータ（次元）での予測値を平均した予測値が、予測の向上を示す事を示した。しかし、考察に述べたように、改善の余地が沢山残されている。さらに多くの数値実験して検討していく必要がある。また他の予測器での予測値との平均等、様々な残された方法がある。

参考文献

- 1)合原一幸編:カオス時系列解析の基礎と応用,産業図書(2000).
- 2)北川源四郎:時系列解析入門,岩波書店(2005).
- 3)E.N.Lorenz:Journal of the atmospheric sciences, 20(1963)130.
- 4)渡辺他:小山高専紀要38(2006)101.
- 5)渡辺:小山高専紀要39(2007)107.
- 6)渡辺:小山高専紀要40(2008)91.
- 7)渡辺:小山高専紀要41(2009)117.
- 8)M.Casdagli:Physica D, 35(1989)335.
- 9)<http://www.fukuoka-jma.go.jp/fukuoka/chosa/taifu/sekkin.html>

小山工業高等専門学校 電子制御工学科
E-mail : watanabe@oyama-ct.ac.jp

「受理年月日 2009年9月29日」

