小山工業高等専門学校研究紀要
第 38 号（2006）101－106

カオス時系列短期予測の予測精度の扱いについて

To the Short Term of Chaotic Time Series Prediction about the Treatment of the Prediction Accuracy

渡辺　達男、長島　栄吾*、関　　拓也*
Tatsuo Watanabe, Eigo Nagashima*, Takuya Seki*

1．序論

近年、カオス時系列の短期予測方法について、幾つかの研究がなされている[1、2]。自然界にはカオス現象と思われるものがたくさん存在するが、そのために自然界における予測が難しくなっている。しかし、最近の研究により、単純なカオス時系列に関しては、その短期予測がある程度可能であることが指摘されている。

渡辺[3、4]は、カオス時系列、特にロジステック写像に関して、Lorenzの類推法[5]を用いて、その予測精度を調べた。Lorenzの類推法では、埋め込み次元の数や、予測に用いる近接点数等により予測精度が変化してくる。渡辺は、近接点数、埋め込み次元数を変化させたとき、予測精度に特異な性質があることを示した。単純に近接点数、次元数を増加させることが予測精度を向上させることにならず、ある近接点数、次元数で最も予測精度を高められることを見いだした。また、予測を行うに用いるデータの数により、最適な近接点数、次元数があることも見いだした。

この論文では、カオス時系列、特にロジステック写像に関する短期予測を議論する。短期予測方法はローレンツの類推法を用いる。昨年度の論文では、データ数の数や、どの点から予測するか、何点先まで予測するか等により、予測精度が著しく変化し、予測精度がどの程度あるのかがあまりはっきりとしなかった。今回の論文では、長島君が考案した方法により、今までよりも予測精度の比較がわかりやすくなってきたので、それを報告したい。

２章は、Lorenzの類推法によるカオス時系列予測の方法、長島君[6]が考案した予測精度の比較方法改善を解説する。

３章は、２章の方法を用いて、実際にロジステック写像の短期予測を行った結果を報告する。

４章は考察に、５章はまとめにあてられる。

2．Lorenzの類推法

LorenzはLorenzの類推法を１９６９年に考案した[5]。これは再構成された状態空間内カオス時系列を考え、現在の点v（t）に対して、その近傍点v'（t'）を検索し、予測点v（t+１）は近傍点v'（t'＋１）で与えられるとするものである。Lorenzはこの方法で、気象データの予測を行ったが、残念ながら成功しなかった。その後この方法は改良され、近傍点を複数取り、

*)小山高専電子制御工学科H1６年度卒業生

それらに適当な重みをつけて、予測点とする方法が取られる。渡辺[4]は内分点を用いて、近傍点から予測点を求め、幾つかの性質を調べた。

$$x_i(t+1)=\frac{\sum_{j=1}^{k} x_{i,j}(t)m_{i,k-j+1}}{\sum_{j=1}^{k} m_{i,j}} \qquad (1)$$

（1）式で、ベクトル$x_i(t+1)$は予測点ベクトル$x_{i,j}(t+1)$は近傍点ベクトルの１ステップ先のｊ番目ベクトル。$x_i(t)$はデータの最後のベクトル。$m_{i,j}$は$x_i(t)$と近傍点$x_{i,j}(t)$との距離を表すベクトル。kは近傍点数である。

（1）式を用いた予測では、埋め込み次元を２、３次元で近傍点数２、３点で予測精度が高いことが示された。

しかし、この方法で予測精度を求めると、ｔが１増加するだけで、真値と予測値の誤差が大きく変化し、予測がどれほど正しいのかが今ひとつ明確では無かった。

長島君[6]は予測精度をより明確にするために、データ数がnの時、n±h（hは１から５０程度）のデータ数に対する同じ予測ステップ数ｓでの誤差を求め、その誤差を平均することにより、データ数nの時の予測ステップ数sの誤差とすることを提案した。これにより、今までは同じ予測ステップ数sでデータ数nが１増えただけで、誤差が大きく変化し、予測精度の性質が見えにくかったが、より精度の性質が見えるようになった。この方法を平均法と仮に名付ける。ここでデータ数とは予測に用いる、既に分かっているデータの個数で、このデータをもとにして、予測をする。いくつ先の時系列データまで予測するかを予測ステップ数sと呼んでいる。データ数nが少なく、予測ステップ数sが多いほど誤差は多くなり、データ数nが多く、予測ステップ数sが少ないほど誤差は少ない。

以下ではLogistic写像、

$$x(t+1)=ax(t)(1-x(t)) \qquad (2)$$

を用いて、埋め込み次元３次元、近傍点数２点を用いて、データ数と誤差の関係等を平均法を用いて見ていくことにする。

３．数値実験

Logistic写像で、データ数n＝５０００個の時、ステップ数と、真値と予想値のグラフを図１に示す。

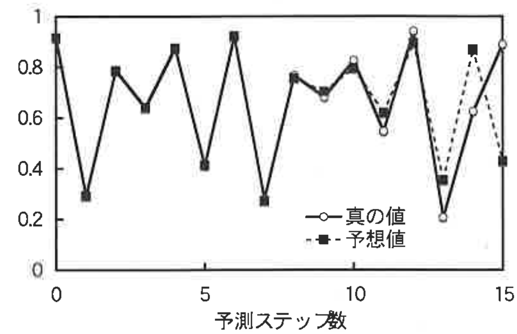

図１　データ数５０００個における、真値と予測値

図からわかるように、予測ステップ数が１０程度までは、良い予想ができている。しかし、１０を超えると、真値と予測値が少しづつズレていき、予測は困難となる。

次に、データ数１０００点付近で１０ステップ先まで予測した時の絶対誤差の関係を図２に示す。

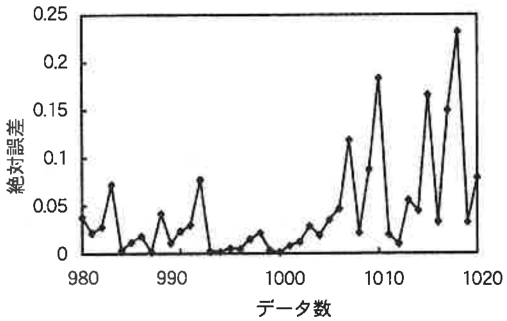

図２　データ数１０００付近での１０ステップ先まで予測したときの絶対誤差の関係。

横軸は予測に用いたデータ数。縦軸は、そのデータで１０ステップ先まで予測した時の真値と予測値との絶対誤差である。この図を見ると分かるが、データ数により、１０ステップ先の誤差が大きく異なる。１０００点付近では誤差が少ないの対して、１０１８点では誤差が大きい。また、その付近では、データ数が１異なるだけで、誤差が大きく異なる。これは、予測をするために、近傍点を探すが、近傍点がどの程度近くにあるかはデータ数により異なるので、このような関係になると思われる。データ数が５０００点付近の１０ステップ先予測における絶対誤差の図を図３に示す。

図を見ると、図２と同様に、データ数により、誤差が大きく異なる。特に、データ数が４０９５点では誤差が大きい。しかし、４９８５点から４９９０点付近、５０１０点付近から５０１５点付近ではあまり誤算に変化が無く、しかも誤差が少ない。これは図２の９９３点から９９６点付近が誤差が少なく、誤差の変化も少ないグラフの形に似ている。

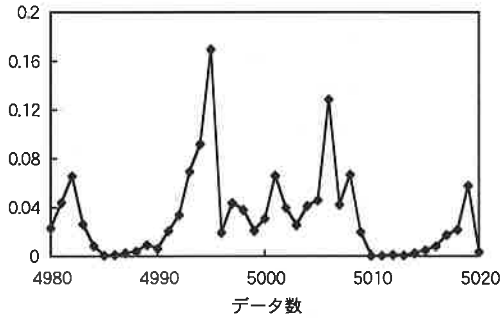

図３　データ数５０００付近での１０ステップ先まで予測したときの絶対誤差の関係。

これらのことから、予測の元となるデータ数が１異なるだけで、大きく誤差が変化する可能性があるので、長島君による次のような平均方法を考える。

すなわち、例えばデータ数１０００点での１０ステップ先の絶対誤差をデータ数１０００点だけでなく、１０００±ｎ点の１０ステップ先の絶対誤差の平均値とするということである。これにより、たまたま誤差が少なかったとか、多かったということが少し除外される可能性が出てくる。ここでｎは５から２０程度に選ぶことにする。あまり小さいと平均を取る必要がないし、大きすぎると何処のデータ数の点を議論しているのか分からなくなる。

この方法でデータ数を１０００点から２０００点まで各データ数の点に対してn＝５点（計１１点）の絶対誤差の平均を取り、関係を調べた。その関係を図４に示す。

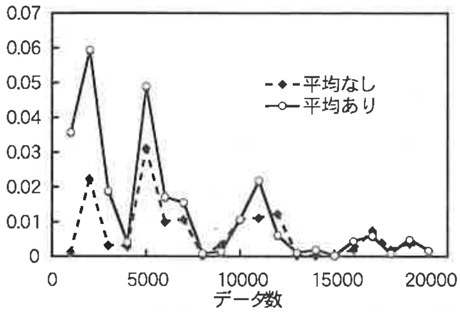

図４　データ数±５個の１０ステップ予測における絶対誤差の平均とデータ数の関係。

図を見ると、データ数が１００００以上では、平均あり、平均なしのグラフにはあまり、差がないが、データ数が３０００以下では、大きく異なる。平均なしではデータ数１０００の誤差は少ないように見えるが、実際は大きいことが見て取れる。またグラフはだいたい右下がりになっており、データ数が多くなれば、予測誤差が少なくなることが見られる。しかし、データ数５０００以下では平均しても誤差の変動が大きい。この程度のデータ数ではやはり、データの個数により誤差が大きく依存することが分かる。

nを増やすとどうであろうか？n＝１０個での同様の関係を図５に示す。

図を見ると、図４とあまり変わらない。nを増やしてもグラフの形に大きな変化がない。従って、平均したグラフが、予測精度の大まかな変化の傾向を表していると思われる。

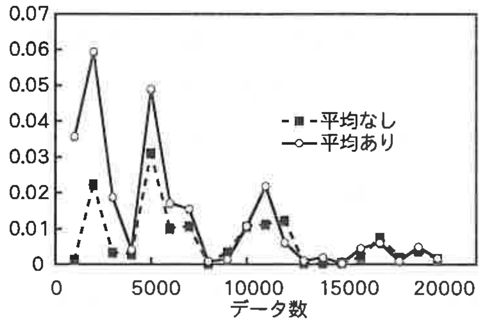

図５　データ数±１０個の１０ステップ予測における絶対誤差の平均とデータ数の関係。

４．考察

３章では、Logistic写像をLorenzの類推法を用いて、予測誤差の平均化による効果を見てきた。

この章では若干の考察をする。

今回報告した平均法を用いると、ある程度データ数が少ないところ（５０００以下程度）では、データ数による誤差の変動が少なくなり、データ数が誤差にどの程度影響を与えているかをある程度見て取れると思われる。

しかし、平均化しても、図４、５でのデータ数４０００点付近でのように、まだ大きな誤差のばらつきがあり、完全にこの方法が成功しているとは言えない。nを増やすことにより、効果がより大きく出ると考えたが、実際にはそれほどでもなかった。これ以上nを増やすと今度は何処の点を考えているかがぼやけてくるので、難しいだろう。

数値実験の結果で、図３を見ると、周期性があるのではないかという仮説を考えることができる。４９８５から４９９０、５０１０から５０１５の範囲はだいたい同じ形をして

いる。周期を計算すると、２５程度になる。予測の方法から何らかの周期的傾向があると考えられるが、周期的に誤差が増減を繰り返し、データ数が増えるに従い、徐々に誤差が少なくなっていくのではないかと思われる。

５．まとめ

Logistic写像に、Lorenzの類推法を用いてカオス時系列の予測問題を議論した。その際、誤差の処理の方法で平均法を提案し、その効果を調べた。その結果、データ数が少ない（５０００以下）では平均化することに意味があることが分かった。しかし、平均化してもまだデータ数により誤差の変動が大きいことも認められた。

今回は簡単な平均法で簡単な検証を行ったが、さらに様々な統計的手法により、よりLorenzの類推法の精度を明らかにすることが考えれる。

また、さらなる精度向上の方法を考えていきたい。

参考文献

１）合原、五百旗：カオス応用システム、朝倉書店(1995)。
２）合原編、池口他：カオス時系列解析の基礎と応用、産業図書(2000)。
３）渡辺：小山高専紀要、36号(2003)137。
４）渡辺：小山高専紀要、37号(2004)179。
５）E,N,Lorenz:J.Atoms.Sci.26(1969)36.
６）長島栄吾：小山高専電子制御工学科卒業研究抄録集（平成１６年度）。

「受理年月日　2005年９月29日」