ニューラルネットによる異常値の検出

正　員　大江　修造†

A Neural Network Methodology for Detection of Abnormal Value

Shuzo OHE†, *Member*

† 東京理科大学工学部経営工学科，東京都

Faculty of Engineering, Science University of Tokyo, Tokyo, 162 Japan

あらまし　ニューラルネットの汎化能力を用いて，数値データの異常値を検出する方法を提案する．関数関係にある2個の変量，独立変数および従属変数において，従属変数としての数値データに異常値が何らかの理由により認められる場合にその検出を行う．

キーワード　ニューラルネット，汎化，バックプロパゲーション，異常値検出

1. まえがき

ニューラルネットにおけるバックプロパゲーション法[1]は関数の連続性を汎化できる能力[2]を有している．この能力を用いて，複数の要因が影響を及ぼす非線形の問題を解明することができる[3]．一方，この汎化能力を用いれば関数の連続値から逸脱する異常値を検出できることが予想される．

異常値の検出はさまざまの分野において必要である．各種の測定において，人間のミスによる場合と，装置の不調による場合とを問わず，通常の測定誤差範囲を超える異常値が測定される可能性が存在する．あるいは，測定値は正常値であるにもかかわらず，測定値の記録や印刷の過程において，誤記・誤植の発生する可能性も存在する．一方，各種の製造プロセスにおいても，製造装置の不具合などにより，温度，圧力，流量などの値に異常値の発生する恐れもある．

これらの異常値の検出は，主に該当する職種の熟練者の経験によるところが大である．しかし，人間を介在せずに異常値を検出できることが望ましい．

関数関係が明確にされていて，式化されている場合であれば，最小2乗法における誤差により異常値を検出できる．しかし，関数関係は認められていても式化されていない場合もある．このような場合に一般化された検出法は存在しない．

そこで，バックプロパゲーション法が関数の連続性を汎化できる能力を用いて，1点の異常値を正常値を学習することなしに検出する方法を提案する．本論文では，このような異常値の検出法の確立を図るために，離散的な数値データを関数式により発生させ，そのうちの1点を異常値と交換した場合につき検出の可能性を述べる．

2. 異常値を検出するアルゴリズム

バックプロパゲーション法における中間層のノード数を多くすることにより，3層のニューラルネットを用いて任意の関数を近似できることが証明されている[4]．この関数近似の能力を用いることにより，関数の連続値からはなはだしく逸脱する誤差の大きな異常値を検出できる．

2.1 アルゴリズム

バックプロパゲーション法を用いて異常値を検出する方法を以下に述べる．

（1）異常値を含む数値の離散データ (x, y) を用意する．

（2）ニューラルネットの入力層に(1)で用意したデータを与える．この際，異常値データは学習させずに異常値以外のデータを用いて学習したニューラルネットに異常値データを呈示して誤差を算出する．いずれのデータが異常値であるか否かは事前には不明であるので，逐一データ中の一点を異常値候補として学習からは除外する．

（3）出力層 (y) の全データに対する誤差が減少し収束するまで学習を繰り返す．

（4）各入力データに対する誤差を求め最大誤差を示すデータを異常値とする．

2.2 離散的な数値データ

離散的な数値データを発生させる関数式として，α を定数とし，$y=-x+1$ につき対称で，発生した数値データの多項式近似が困難である次式を用いた．

$$y=\frac{\alpha x}{1+(\alpha-1)x} \tag{1}$$

2.3 ニューラルネットの構造

異常値の検出に用いたニューラルネットの構造は，入力層・中間層・出力層の3層からなり，入，出力層のノード数は1である．入力層には用意されているデータ (x) を入力し，出力層からは (y) を得る．教師データとしては用意されているデータの y の値を使う．中間層のノード数は，あらかじめ正常値のみのデータにより最適なノード数を決定しておいた．バックプロパゲーション法における学習効率および過去学習に関する定数は，0.7および0.8とした．

3. 計算機実験

3.1 異常値を含むデータ

異常値を検出するために，式(1)における α の値を2.5

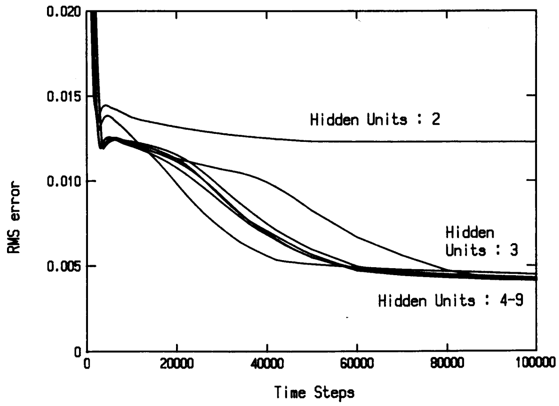

図1 ニューラルネットにおける学習の挙動曲線
Fig. 1 Network Performances curves for learning.

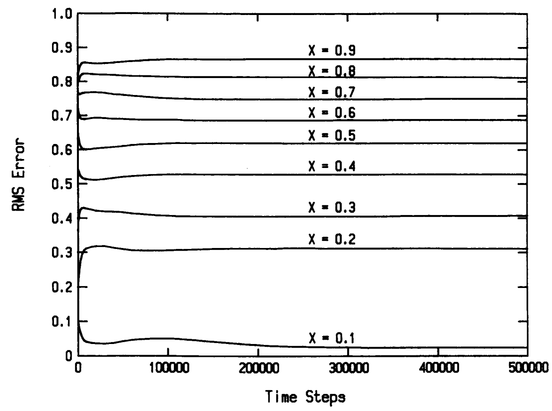

図3 各 x の値における異常値 $y=0.1$ に対する誤差
Fig. 3 Errors for abnormal value of $y=0.1$ at each x.

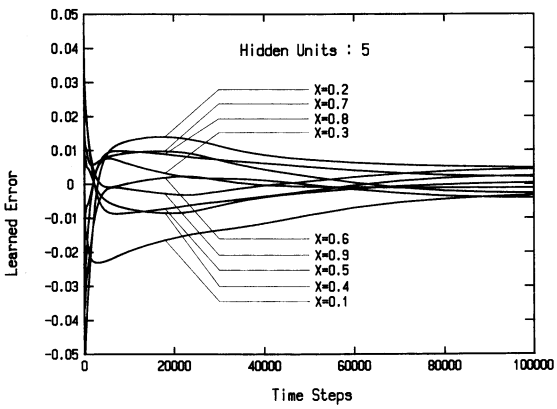

図2 式(1)の各 x の値に対する学習の挙動曲線
Fig. 2 Learning Performances curves for each x values of Eq. (1).

とし，x の値を0から1まで0.1間隔ずつ増加させて y の値を求めた．各 x の値に対する y の異常値として0.1を採用した．すなわち，x の値0.1における正常値 $y=0.217$ に対して異常値として $y=0.1$ を採用して，その他の x の値に対しては式(1)の y の値を正常値として採用する．以下同様にして，特定の x の値に対して y の異常値を0.1とし他の x に対しては，式(1)の y の値を正常値として採用する．このようにして，x の値を0.1から0.9までの9組の異常値のデータを準備する．

3.2 中間層におけるノード数の決定

2.3 で述べたニューラルネットにおいて，中間層におけるノード数を2から9まで変えて，式(1)の正常値により最適なノード数を決定する．各ノード数に対する学習誤差を図1に示した．図1からわかるように学習回数80,000回で，ほぼ収束しているので，学習回数としては100,000回とした．学習誤差はRMSで表した．ノード数3以上ではそれほど学習誤差に差はない．図1からノード数の増加と共に学習誤差は減少していることが認められるがノード数4以上ではその差は小さい．ニューラルネットの学習に要する演算時間を考慮すれば，ノード数は少ない方がよい．従って，中間層のノード数としては5を採用する．

中間層のノード数を5とした場合の x の各値に対するニューラルネットの学習状況を図2に示した．図1における学習誤差はRMSで示したが，図2では，学習誤差は各 x に対する誤差で示した．$x=0.1$ の場合における収束には回数を要しているが，$x=0.1$ 以外の場合は，80,000回で既にかなり収束している．

3.3 異常値の検出

ニューラルネットにおけるバックプロパゲーション法は関数の連続性を汎化できるから，関数関係から離脱する異常値は大幅に大きな学習誤差を示す．図3に異常値の学習誤差を示し，図4～6には，それぞれ $x=0.8$，$x=0.6$，$x=0.5$ における正常値と異常値との学習誤差を併せて示した．異常値を検出するための学習回数は図3～6に示したように500,000回とした．異常値を含まない式(1)の学習については100,000回で十分であることが図1，2からわかるが，異常値を含むデータの検出には100,000回では不十分であることが図4～6からわかる．もっとも，図4～6から学習回数100,000回によっても異常値の検出は可能ではある．例えば，図4からわかるように学習回数100,000回において $x=0.8$ における異常値に対する誤差は0.814であるのに対

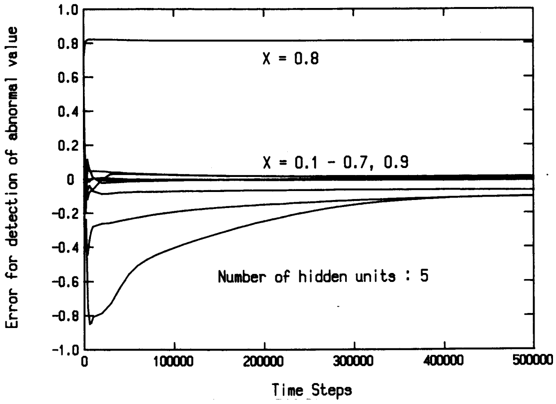

図4　学習誤差による $x=0.8$ における異常値 y の検出
Fig. 4　Detection of abnormal value of y on learned error at $x=0.8$.

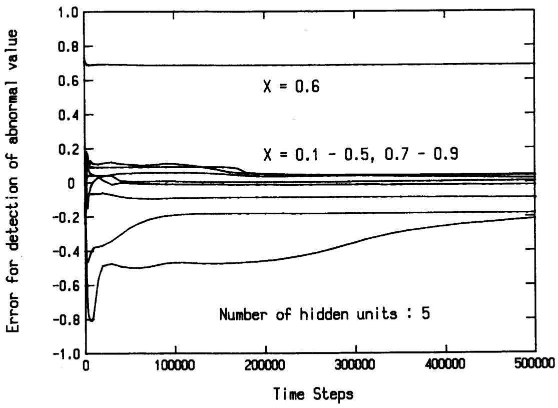

図5　学習誤差による $x=0.6$ における異常値 y の検出
Fig. 5　Detection of abnormal value of y on learned error at $x=0.6$.

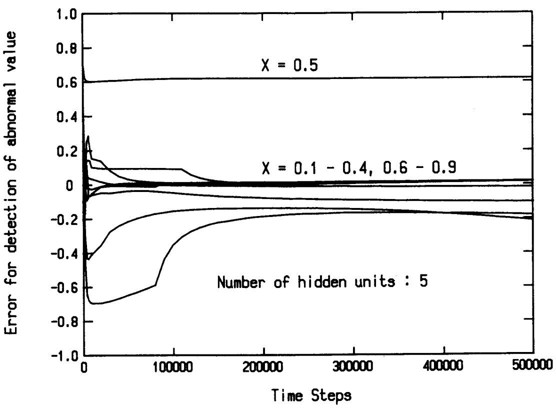

図6　学習誤差による $x=0.5$ における異常値 y の検出
Fig. 6　Detection of abnormal value of y on learned error at $x=0.5$.

表1　中間層のノード数と異常値の検出能力

中間層のノード数	誤　差
5	0.747
6	0.748
7	0.737
8	0.712
9	0.710

して正常値に対する $x=0.9$ におけるそれは -0.405 となっていて正常値に対する x の中では大きい（絶対値）が，異常値における 0.814 の約 1/2 程度と小さい．しかし，図 4 からわかるように学習回数を 500,000 回とすることによって，正常値に対する $x=0.9$ における誤差は -0.104 となり他の正常値の x に対する値と同じ程度になっている．図 3 における各 x に対する RMS は 0.024 から 0.867 に及んでいる．これらの RMS の値は正常な場合に比較して大幅に大きな値であり，これによって異常値を検出できる．誤差の値は x の増加と共に増加している．これは異常値として各 x に対して共通の $y=0.1$ を用いているために，x が増加するに従い異常の程度が増大していることによるためである．

中間層のノード数と異常値の検出能力の相関を表 1 に示した．表 1 は，$x=0.8$ において y の値を異常値 0.1 として，中間層のノード数を 5 から 9 まで変えた場合である．学習回数は 500,000 回である．表 1 において誤差は $x=0.8$ における誤差の値から $x=0.1$ における誤差の値を減じた値である．この誤差の値は異常値の検出能力を示す．表 1 から中間層のノード数の増加と共に誤差の値も減少していることが認められる．

3.4　計算機システム

バックプロパゲーション法のプログラムを FORTRAN により記述し，LAN 上のワークステーションを用いて計算した．

4.　む　す　び

数値データの異常値をニューラルネットにおけるバックプロパゲーション法が関数の連続性を汎化できる能力を用いて検出することができた．

これによって，各種の測定における通常の測定誤差範囲を超える異常値，測定値の記録や印刷の過程における誤記・誤植および各種の製造プロセスにおける温度，圧力，流量などの異常を人間を介在せずに，ニューラルネットにより異常値を検出できる可能性を示すことができた．

今後の課題として，さまざまな異常値についてその

検出におけるニューラルネットの挙動を検討する必要がある．更に，関数式によらない離散的なデータにおける異常値の検出についての検討も必要である．

中間層のノード数と関数近似の能力の関係は明らかではない．よって，種々の関数形について，中間層のノード数と関数近似の能力との関係を計算機実験によって検討しておく意味がある．この点で，本論文は関数形を限定しているが，ニューラルネットによる異常値の検出法の確立に有効である．

文献

(1) Rumelhart D. E., McClell J. L. and PDP Research Group : "Parallel distributed processing", vol. 1 and 2, MIT Press, Cambridge, MA. (1986).

(2) 上坂吉則："ニューラルネットと学習可能性"，信学誌，**74**, 9, pp. 943-948 (1991).

(3) 大江修造："A neural network prediction system for critical temperature of paraffin"，石油学会誌，**35**, 1, pp. 107-110 (1992).

(4) Funahashi K. : "On the approximate realization of continuous mappings by neural networks", Neural Networks, 2, pp. 183-192 (1989).

(平成5年8月19日受付)