

気象の予測と研究のためのニューラルネットワーク I*

柳野 健**

1. はじめに

ニューラルネットワークは気象予測や気象研究にとって大変有用であるが、そのネーミングは誤解を与えやすい。人工知能のような、神秘的な自由意志を連想させる。筆者は平成8年3月からスタートした新しい天気予報に関連して、ニューラルネットワークによるガイダンスの開発に関わっていた。その頃、予報官を対象とした研修において、このテーマで何度か講師を務めたが、講義後の受講者の反応は「ニューラルネットワークを全く誤解していた」というものであった。

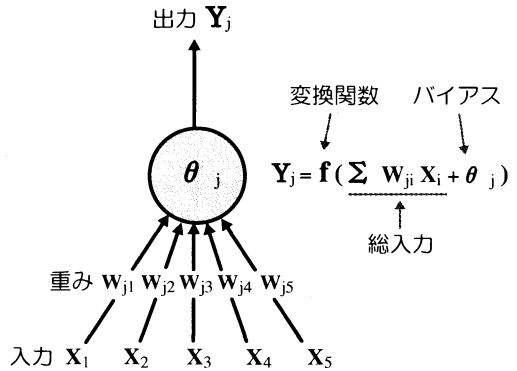
本稿の目的は、ニューラルネットワークを気象の予測と研究に活用するとき役立つ基礎知識を図解で提供することにある。今回は主に学習の仕組みを、次回は気象への応用を述べる。

2. ニューラルネットワークの定義

本稿では、3層の階層型ニューラルネットワークのみを扱う。最も基本的であり、最も活用されている。

生物の神経の基本単位はニューロンである。神経細胞本体、本体周辺部の樹状突起、本体から伸びた神経繊維からなる。樹状突起で他のニューロンからの信号を受け取り、神経細胞本体で総合処理し、処理結果を神経繊維を通じて他のニューロンへ信号として出す。

そのはたらきを単純化したのが、第1図に示す数学モデルである。入力端子は複数あり、ニューロン本体では、各々の入力値に重みを掛けて和をとった総入力に、バイアスを加え、変換して出力する。出力端子も複数ある。どれも同じ値を出力するので、モデル図では、1本の出力端子だけを代表して描く。変換関数は



第1図 人工ニューロンの数学モデル。

重要で、素性の良いS字型のシグモイド関数(第5図)がよく使われている。もちろん、他の関数であってもかまわない。ガボール関数(後述)なども使われる。

気象現象などは、ある臨界値を超えると急激な変化が発生し、やがて飽和する。そのような変化の全域を無理なく表現できる。それが、人気の理由であろう。飽和も含めた全域表現は重要である。バブル経済絶頂期に、地価の高騰をある等比級数で表わす理論を説く経済学者がいた。その理論では地価は無限に高騰する。

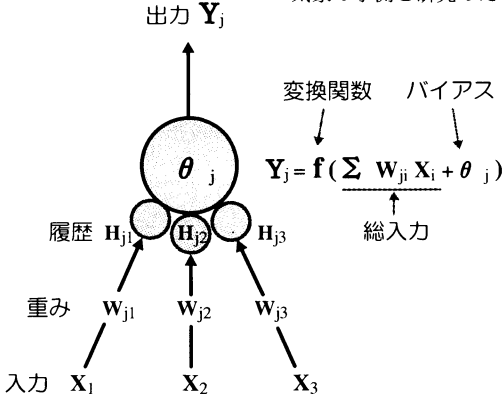
神経系の研究が進めば、ニューロンのモデルもまた進化する。第2図は、最新のニューロンモデルである。過去の入力の痕跡を保持しておき、今、電気ショックのような強い入力に対応した出力があると、過去の入力の痕跡を調べ、ある値以上ならば、電気ショックとその入力との結合を強める。こうして、時間的に継続して起こる事柄を関係づける学習ができるようになる。

ここでは、標準型ニューロンを結合する。第3図は、最も簡単な結合例である。ニューロンは入力部でも出力部でも複数個あるのがふつうである。入力を受けつ

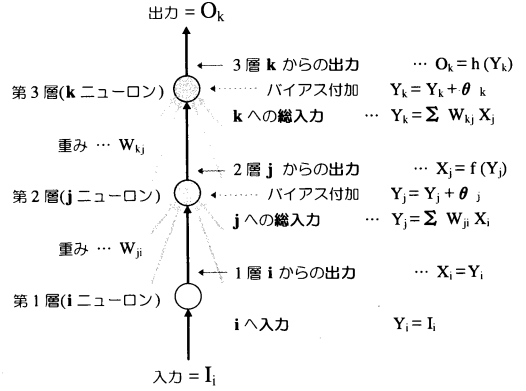
* Neural Network for Forecast and Research in Meteorological Phenomena Part I.

** Takeshi Yanagino, 気象大学校.

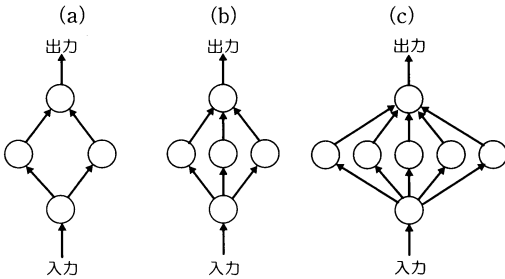
© 1998 日本気象学会



第2図 人工ニューロンの最新モデル.



第4図 ニューロンの結合と、入力情報の順方向への伝播.



第3図 階層型ニューラルネットワーク. 入力層で1, 中間層でM, 出力層で1個のニューロンを持つタイプ(階層型1-M-1). (a) 階層型1-2-1, (b) 階層型1-3-1, (c) 階層型1-5-1.

けるニューロンを入力層, 出力を受けもつニューロンを出力層, その中間のニューロンを中間層と呼ぶ.

具体的に, それぞれの層のニューロンの働きを見てみよう(第4図). 入力層のニューロンは, 他と異なる. 入力1つで, そのままの値を中間層の各ニューロンへ出力する. 中間層のニューロンはそれぞれの入力値を重みづけして総和をとり, バイアスを加えて, シグモイド関数 f を通して, 出力層の各ニューロンへ出力する. 出力層のニューロンはそれぞれの入力値を重みづけして総和をとり, バイアスを加えて, 出力関数 h を通して出力する. h は f と同じである必要はない. 通常は, h もシグモイド関数としている.

3. ニューラルネットワークの仕組み

さて, 階層型ニューラルネットワークの仕組みを調べよう. そのため一番簡単なものを使う(第3図a). これは, ニューロンが入力層で1個, 中間層で2個,

出力層で1個なので, 階層型1-2-1と略記しよう.

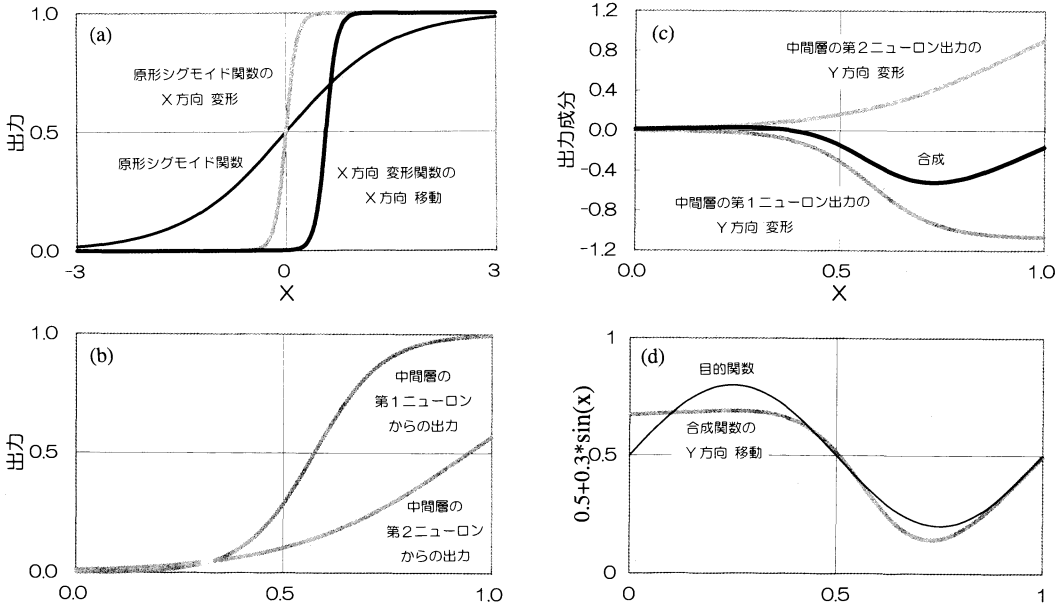
階層型1-M-1は, 1変数の未知関数を学習できる. $y=f(x)$ だが, 関数 f は未知である. ある x_1 のとき y_1 で, 別の x_2 のとき y_2 でと... そのように, データが万遍なく与えられれば, ほぼ全域の関数関係を学習できるから, f はこんなものだろうと, 帰納的に構成するのである.

階層型 N-M-1は, N変数の未知関数を学習する.

原理的には, M (中間層のニューロン数) が大きいほど, 複雑な関数を表現可能である. しかし, 扱う問題ごとに, 表現したい関数関係の複雑さは異なる. そのような訳で, 適当な M の数も扱う問題ごとに異なる.

入力層は入力値を取り次ぐだけである. 中間層の2つのニューロンへ同じ入力値を伝える. ふつう, この入力値は (0~1の大きさに) 規格化している.

中間層のニューロンは何をするのだろうか? 中間層のニューロンは, 変換出力用に同じシグモイド関数(原形)をもっている(第5図a). 中間層のニューロンへの入力値 (X) を重みづけ (W 倍) することの意味は, そのニューロンの原形シグモイド関数を, X 軸の原点を中心にして, X 軸方向に圧縮・伸長・反転することである(第5図a). W が1より大きい場合は圧縮で, 1より小さい正の場合は伸張である. W が負の場合は反転した圧縮・伸長である. バイアスを加えることの意味は, 圧縮・伸長・反転等で変形されたシグモイド関数を X 軸方向にずらすことである(第5図a). バイアス値が負であれば右, 正であれば左へずらす. 第5図bは, 2つの中間層のニューロンからの出力である.



第5図 階層型ニューラルネットワークの仕組みの解明(階層型1-2-1の場合)。(a) 中間層のニューロンは、原形関数を X 方向で変形・移動する。(b) 中間層のニューロンから出力される、X 方向での変形関数。(c) 出力層のニューロンは、(b) を Y 方向で変形・合成する。(d) 出力層のニューロンは、(c) を Y 方向で移動し、出力する。

出力層のニューロンは何をするのか？ 筆者は出力層のニューロンの変換関数を線形関数としている。理由は学習が高速になるためであるが、その働きも理解しやすくなる。この出力層のニューロンへの入力値を重みづけ (W 倍) することの意味は、中間層の個々のニューロンで、 X 軸方向の変形を受けたシグモイド関数を、 Y 軸の原点を中心にして、 Y 軸方向に圧縮・伸長・反転することである(第5図c)。総和をとることは、それらを合成することである。バイアスを加えることは、そのような合成関数を Y 軸方向にずらすことである。第5図dの太線は出力層ニューロンからの出力結果である。細線はある SIN 関数である。実は、この例は階層型1-2-1で、その SIN 関数を未知関数として、実際に学習させた結果である。

4. 中間層のニューロン数の意味と実例

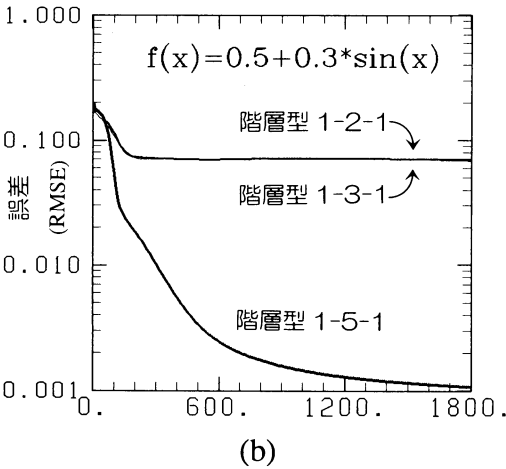
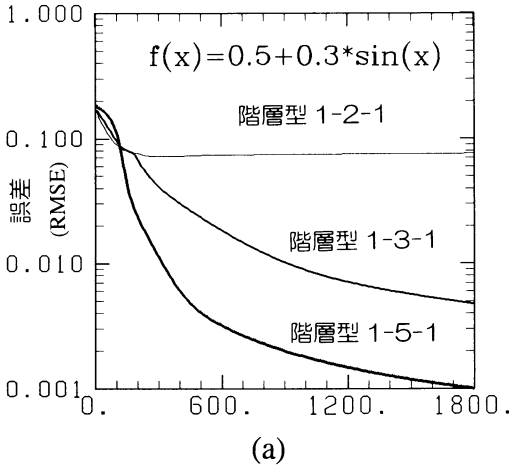
前節をまとめると、ニューラルネットワークにおける学習とは、知りたい未知関数を最も良く近似するために、中間層のニューロン数だけあるシグモイド関数を、適切に変形・移動・合成するような「重みやバイアス」を求めることである。この節では実例によって、さらに理解を深める。

フーリエ解析と比べて面白いのは、複数の原形関数(シグモイド関数)を少しずつ変形して合成するという柔軟なやり方をしていることである。つまり、逐次解法による関数展開システムである。驚くべきことは、互いに直交しない関数系(シグモイド関数など)を用いて、任意の関数を、それなりに表現する能力である。

もう少し正確に言えば、任意の関数を、望む精度で近似表現するような、シグモイド関数による3層ニューラルネットワークは存在する。逐次解法(学習法とも言う)でそのような解へ漸近させる。山頂への登り方が、ただ1通りではないように、さまざまな逐次解法(学習法)がある。

では、中間層のニューロン数の意味を理解するために、具体的な実験をしよう。 $f(x) = 0.5 + 0.3 * \sin(x)$ を使い、 (X_i, Y_i) の組データを21個用意して、階層型1-2-1, 1-3-1, 1-5-1でそれぞれ学習させたとき、学習の進行につれ、この未知関数と合成関数との誤差(RMSE=平均自乗根誤差)はどうなるだろうか？

ここで未知関数という言葉の不審に思う方のために補足する。ニューラルネットワークにとっては、学習は常に未知関数との遭遇である。ある変数値 X_i のとき Y_i が対応したという個々のデータ対を与えられ、その

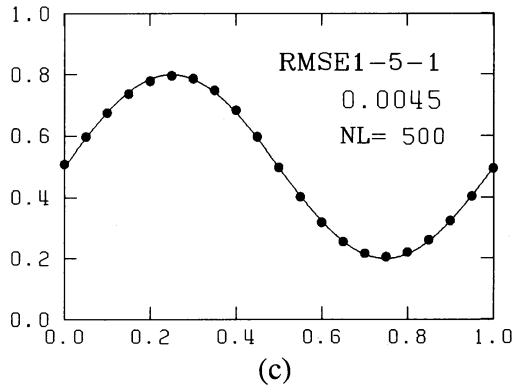
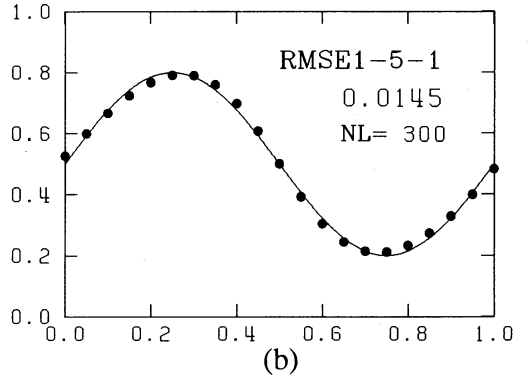
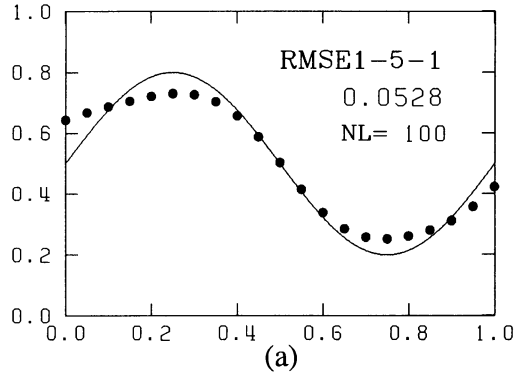


第6図 階層型1-2-1, 1-3-1, 1-5-1の学習の進み具合. $f(x) = 0.5 + 0.3 \sin(x)$ を学習させた. (a)階層型1-3-1, 1-5-1共に学習に成功する. (b)階層型1-3-1は, 学習に失敗する.

背後にある未知関数を構成させられるからである.

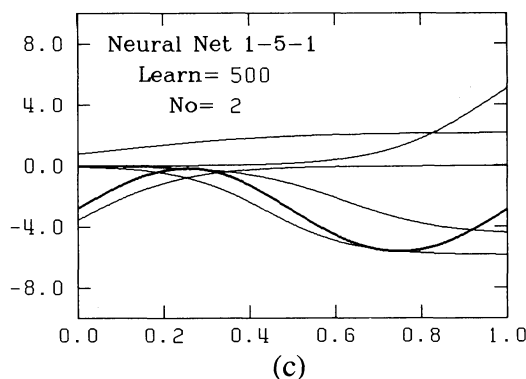
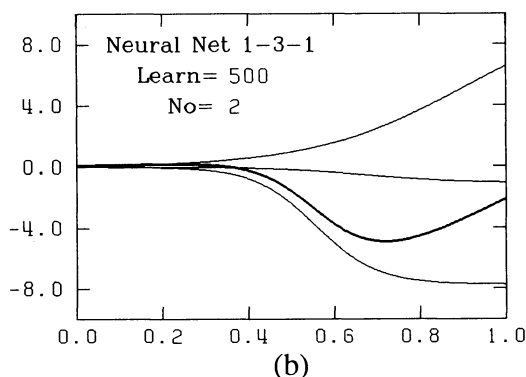
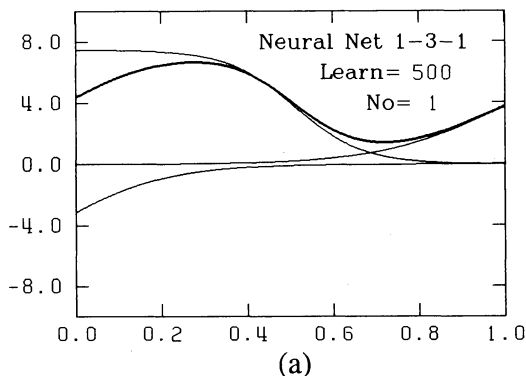
この節では, われわれにとっては既知の関数で学習データ対 $(X_i, f(X_i); i=1 \sim 21)$ をつくる. 正解が分かっているから, どのくらいの学習能力かを評価できる.

第6図aの細線, 中線, 太線は階層型1-2-1, 1-3-1, 1-5-1に対応する. 階層型1-2-1では途中で誤差が減らなくなる. その理由は前節のことから分かる. 中間層のニューロンが2個だから, 2つのシグモイド関数を変形合成してSIN関数一周期分を近似せよということになる. 残念ながらこれはできない. ノートに描い



第7図 階層型1-5-1, 誤差(RMSE)と近似程度. $f(x) = 0.5 + 0.3 \sin(x)$ を学習させた. (a)学習100回, (b)学習300回, (c)学習500回.

て考えて欲しい. 許される変形は, 左右への圧縮・伸張・反転と移動, また上下への同様な変形と移動である. 変形後もシグモイド関数の範疇にある. 2つのシグモイド関数を使った変形合成では, 1山分の変化, SIN関数ならば, 半周期分の表現しかできない. 3つ



第8図 階層型1-M-1, 学習内容 $f(x) = 0.5 + 0.3 \sin(x)$ を学習させた. (a) 階層型1-3-1学習に成功する, (b) 階層型1-3-1学習に失敗する, (c) 階層型1-5-1学習に余裕あり.

のシグモイド関数が使えれば, 2山分の変化, SIN関数ならば, 一周期分を表現することも可能になる.

だから, 階層型1-3-1は近似に成功した. 階層型1-5-1はそれよりも素早く誤差が減少していく.

ここで, 誤差の大きさと近似程度を知ってもらうために, 第7図に目的関数へ近似関数が近づいていく様子を示す (NLは学習回数). 通常, 目的関数も規格化する. RMSE=0.01程度は実用に耐える精度であることが, 図から納得できるだろう.

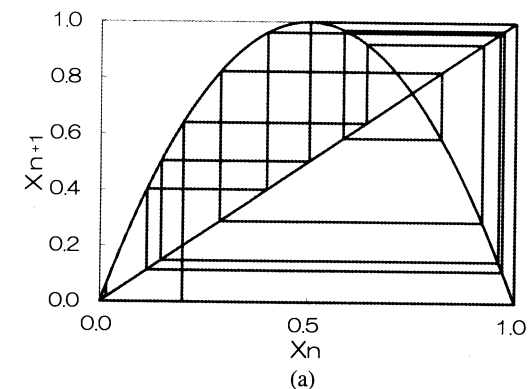
第6図bは, 階層型1-3-1で近似に失敗した例である (RMSE=0.4で行き詰まり). なぜそうなるのだろうか? その過程を前節の観点から見たのが第8図である. 第8図aは成功例である. 3つのシグモイド関数の変形・移動は, 役割分担を果たし, その合成はSIN関数をうまく近似している. 第8図bでは1つのニューロンは, 自分の役割を果たさず, 他の変形シグモイド関数の垂流になっている. これは目的のSIN関数の右半分を近似するという, 手近な成功への袋小路に入り抜け出せなくなった状態である. 袋小路をローカル・ミニマムとも呼ぶ. 第8図cでは, 階層型1-5-1で近似が余裕をもってよくなっていく. この場合, 中間層のニューロン数に2個の余裕があるためである.

中間層のニューロン数が余裕なしの場合, ローカル・ミニマムに陥りやすいことは納得できただろう. では, なぜそうなるのだろうか? ニューラルネットワークでは, 学習を始める前に, 個々の重みやバイアスとして, 小さい乱数を割りあて, 個性を持たせる. 「梅檀は双葉より芳しい」の通り, 偶然与えられた乱数は, 変形合成の小さな第一歩になっている. それぞれの最初の一步が最終的な変形合成への途上にあるとは限らない. 上記の場合でいえば, SIN関数半周期分を近似目標にするのに大変適した初期状態だったとすれば, そのように最適化し始める. しかし, その目標を達成 (部分最適化) した後は行き詰まることになる.

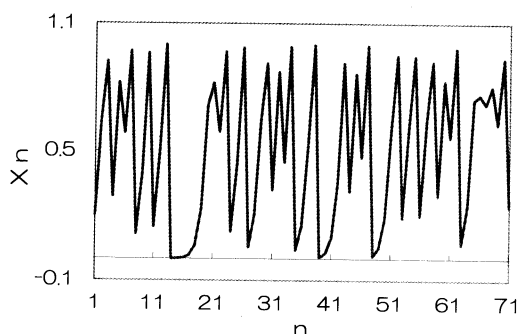
そうなったとき, ランダムに重み等を変えて突破口を開こうとする, 試行錯誤型の学習法がある.

また中間層のニューロンの働きをモニターしていて, 不足状態になれば新たなニューロンを補給し, 働きの悪くなったニューロンは削除する学習法もある. 合理的である. 研究開発型組織のドライな運営に似ている.

ふつうは, やや余裕のあるニューロン数を用意して, 初めは遊軍だったニューロンが, 出番が来れば水を得た魚のように働き始めるようにする. つまり, 全ニューロンが働き過ぎて現状に過剰適応しないように, むしろブレーキをかけておく方法がとられる. その真の意味は, 新しい事態にも対応できるように遊軍を確保しておくことであろう. とても示唆的である.



(a)



(b)

第9図 カオス的な変動。(a) $X_{n+1}=4X_n(1-X_n)$ となる写像, (b) その時系列. 不規則に変動する.

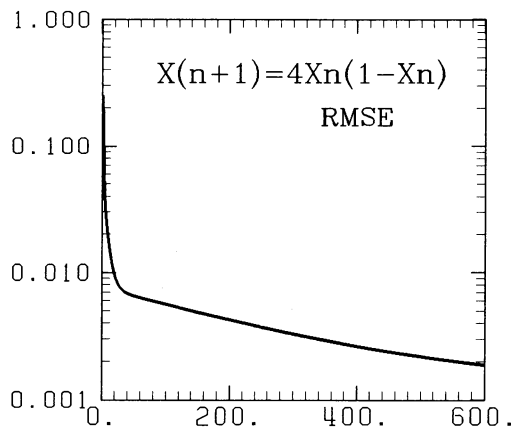
5. カオスの学習

以上で, 任意の関数関係をそれなりに表現できそうだと納得してもらえらるだろう. ところで, 関数関係にはいろいろある. 時間経過における関数関係もある. 系の時間発展とか呼ばれている. このようなものも学習でき, 3時間先までの予測(短時間予測)などに有効である. この節では時間発展を学習する例をあげる.

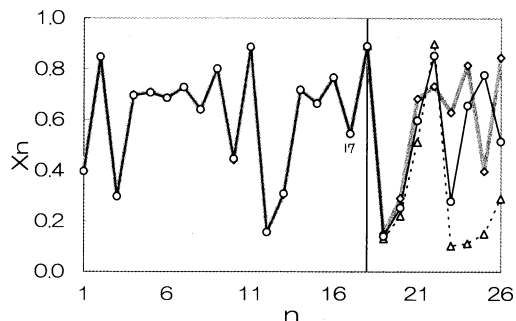
第9図は, 単純ではあるが面白い非周期的な変動である. ある時刻 t_n での変数値が X_n であると, 次の時刻 t_{n+1} での変数値 X_{n+1} は, $X_{n+1}=4X_n(1-X_n)$ と決まる. 繰り返しているうちに, 初期の小さな差異が, やがて大きな違いになってくる(初期値敏感性).

第10図は, このようなデータ対 (X_n, X_{n+1}) を学習させたとき誤差が減少する様子である. $X_{n+1}=f(X_n)$ の f は, 非線形関数である. X_n を現在とすると X_{n+1} は, 1ステップ未来である. だから, この次はどうかという, 時間発展の学習になっている.

第11図の時刻17までは, 学習に用いたデータである.



第10図 階層型1-5-1で, 第9図のデータ (X_n, X_{n+1}) を学習したときの, 学習の進み具合.



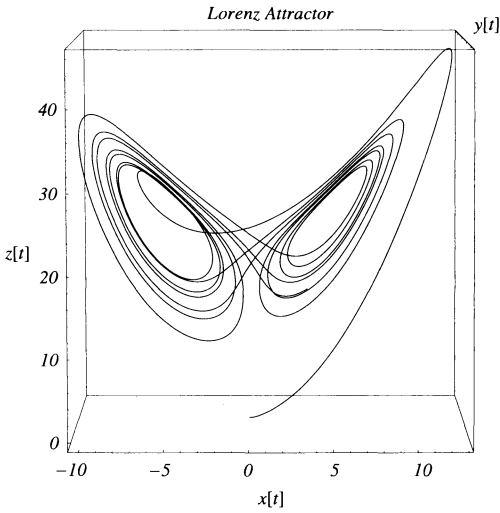
第11図 階層型1-5-1で, 第10図のカオス的な変動を学習(データ1~17)し, 延長予測(データ18~26)した結果.

- ① ○—○ 正しい時系列変化.
- ② ◇—◇ 階層型1-5-1による学習と延長予測.
- ③ △---△ 微小誤差を入れ, 正しい法則で延長予測.

正しい変動は○印と細線, 太線はニューラルネットワークによる計算値である. その誤差は大変小さい.

1ステップ未来はどうかの規則を十分に学習した.

レーダー・エコーを5分間隔で測定している場合, 1ステップ未来とは, 5分後のことである. しかし, 2時間先程度の未来の予測にしないと実用的な価値がない. どうするか? 1ステップ未来は予測できる. その予測通りになったとしたら, 次の1ステップ未来も予測できる. そのように, 1ステップずつの予測を積み重ねていけばよい. すると, 1ステップ未来予測の精度が高い程, 予測を重ねたときの狂いも少なくなる



第12図 E. ローレンツのアトラクター. 対流現象のカオスを表わす.

だ.

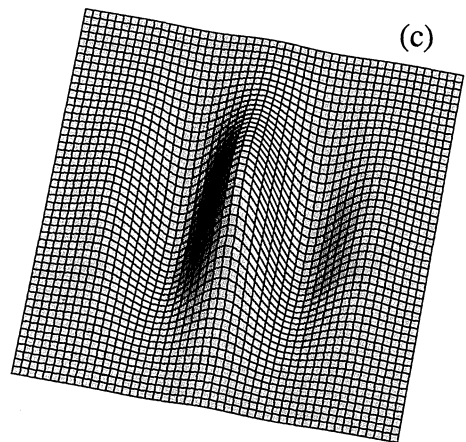
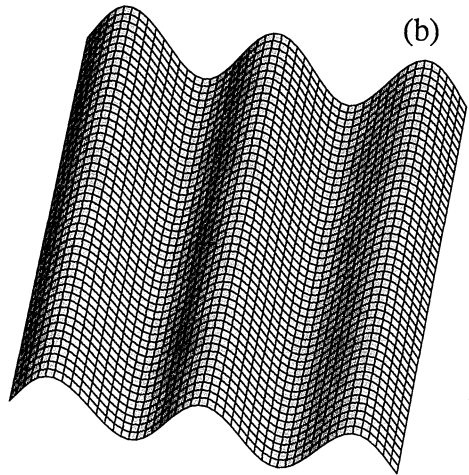
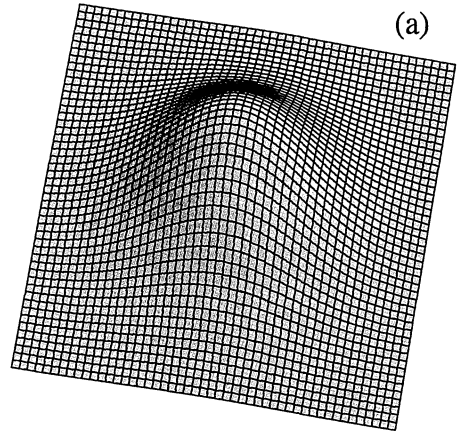
第11図の時刻18からは、1ステップ未来予測を重ねていった予測延長の例である。このときニューラルネットワークの1ステップ未来予測の精度は高いのだが、4ステップ先の予測までが限度である。点線の細線は、時刻18で0.00466の誤差を入れ、その後は完全に正しい法則で未来予測した例である。これも4ステップ先の予測までが限度である。この例はカオスと呼ばれ、長い予測延長が特に困難なものである。

第12図は、対流現象の本質を表現したローレンツ・アトラクタである。ときどき対流状態がガラリと変わることを示している。そのため、初めの僅かの違いがある時点から大きな違いになってしまう。気象現象は多かれ少なかれカオ斯的性格をもっている。数 km 程度のメッシュで降水セルを未来予測する場合は、2～3時間先あたりから急速に難しくなってくる。

6. ガボール関数展開によるメソ気象解析の可能性

D. ガボールは、空間的なひろがり Δx と周波数のひろがり Δf とは相反関係にあることを指摘し、双方の積 $\Delta x \cdot \Delta f$ を最小にするガボール関数を提唱した。人間の視覚も神経レベルでこの関数が深く関わっているようである。

第13図は、2次元ガボール関数の例である。お椀を伏せたようなガウス関数と正弦波の積によってできる局所的な波形である。この2次元ガボール関数 $g(x,$



第13図 ガボール関数. 特徴抽出に適している。(a) ガウス関数, (b) 正弦関数, (c) ガボール関数.

y)を、シグモイド関数と取り替えたニューラルネットワークは、2次元の気象場 $z(x, y)$ を近似できる。長所は、前線や不連続線のような局所的に変動の大きい部分を、少数のガボール関数で表現できる。つまり、メソスケール気象場の重要な特徴を、コンパクトに抽出(情報圧縮)できることである。人間はものごとを特徴づけることによって情報を圧縮し、ささやかな一時記憶容量を実に効果的に運用している。

得られるものは展開係数であり、抽象的な量である。しかし、この展開係数と顕著な気象現象とを対応づける作業もニューラルネットワークにさせることができる。例えば、紙幣の判別の場合、紙幣の画像を2次元のフーリエ級数に展開して、ニューラルネットワークに、その展開係数の特徴と紙幣の金額を対応づけさせることで高い判別精度を得ることができる。ただし、

前線のような急変する場を表現するには、2次元ガボール関数の方がより適しているという訳である。

対応づけられた展開係数から、気象場を再構成できる。主要な展開係数だけで再構成すると、特徴が浮かび上がる。全ての展開係数で再構成すると、ほぼ完全な復元になる。

これはメソスケール現象解析のように、特徴を抽出し、対応づけている人間の理解の仕方に近いやり方を計算機上で実行するという話である。

大量のデータを処理すると、ある現象に対応する気象場の特徴が、逆に、気象場の特徴に対応する現象が、頻度順に整理されて出てくるようにすることも可能である。絞り込まれた特徴から、論理を組み立てたり、考えを洗練したり、現象理解に利用できるだろう。

「陸域水収支に関する国際ワークショップ」のお知らせ

International Workshop on Land-Surface Water Budget

— Toward Hydrometeorology in the Next Century —

Satellites, 4DDA and International Cooperation

気象研究所は平成10年度科学技術振興調整費により、標記国際ワークショップを平成11年1月つくば市で地球フロンティア研究システムと共に開催いたします。

近年、気候変動において陸面過程の果たす役割の重要性が認識されるとともに、今年に GAME の集中観測 (IOP) が実施され、アジア大陸の各地域で陸面水循環過程の観測が行われてこの分野の研究が大きな関心を集めております。

このような折、陸域水循環に関して今後推進すべき研究課題(衛星観測、4次元同化等)、データ交換及び国際協力体制の確立などについて、国内外の関連研究者及びこの分野に関心をお持ちの参加者による意見交換・提言を目的として、本ワークショップを以下の要領で開催致します。

タイトル：陸域水収支に関する国際ワークショップ

主催：気象研究所、地球フロンティア研究システム

日時：1999年1月19日(火)～21日(木)

場所：つくば市第1ホテル

形式：発表は招待者に限られますが、参加は自由(無料)です。

講演要旨集は当日配布致します。

テーマ：

- ① 衛星観測の利用
- ② 陸域4次元同化
- ③ 国際協力体制
- ④ モデリング・季節予報へのインパクト

事務局：気象研究所気候研究部第5研究室

高橋 清利

e-mail : ktakahas@mri-jma.go.jp

Tel. : 0298-53-8670

Fax. : 0298-55-2683

http://gain-hub.mri-jma.go.jp/workshop.html