実数値GAを用いたデータマイニングによる ニューロ降雨予測システムの設計

学生員 伊藤 征嗣* 正員 満倉 靖恵*
正員 福見 稔* 非会員 赤松 則男*

Neuro Rainfall Forecast with Data Mining by Real-Coded Genetical Preprocessing

Seiji Ito*, Student Member, Yasue Mitsukura*, Member, Minoru Fukumi*, Member,
Norio Akamatsu*, Non-member

In this paper, rainfall is predicted by using a Neural Network(NN) and a Genetic Algorithm(GA). GA selects data needed to predict the rainfall. NN learns and predicts it using attributes selected by GA. The real-coded GA is used to decide data priority, and data really needed for the rainfall forecast are selected based on the priority. In order to show the effectiveness of the proposed rainfall prediction system, computer simulations are performed for real weather data. Finally, the effectiveness of this system is shown with data analysis.

キーワード：ニューラルネットワーク, 実数値 GA, 降雨予測

Keywords: Neural Network, Real-coded GA, Rainfall Forecast

1. 緒言

我々の日常生活は、意識するしないに関わらず、多くの場合において天気と深く関係している。したがって、天気を予測するという事は重要な意味を持つ。また、自然災害は異常気象が原因であるものが多く、天気予報を予測することによってその被害を最小限に抑えられることが可能になる。そこで近年、気象予報士という国家資格の設立や、コンピュータの発達などによって、天気予報は序々に予測精度が改善されつつある。現在行なわれている予測方法は、スーパーコンピュータを用いてあらゆるデータから数値的に予報を行い、天気図を作成している。これをもとに気象予報士が知識や経験を生かして一般の人々に分かりやすい予報を発表している。しかしながら、実際の天気予報においても、予測が外れる場合も少なくない。具体的には、次の日における天気予報の的中率は約 80%であるといわれている⁽¹⁾。これは予報士にとっても、また、日常生活においても重大な問題である。天気予報の精度を上げるためには、地域的情報(観測データを含む)を考慮することも重要であると考えられるが⁽²⁾⁽³⁾、実用化には至っていない。ところで、昔からの言い伝えとして、「夕焼け空には翌日に晴れる

兆し」などがあり、比較的的中することは我々自身が日頃感じることである(観天望気という)。これは局所的地域から局所的地域の気象予測が可能であることを示唆している⁽¹²⁾。そこで、本論文では局所的データを用いて局所的な地域の降雨予測を行なう。降雨を予測する理由は、気象予報の中でも一番重要であり、我々の生活に大きな影響があるためである。また、前に述べた観天望気の例でもあるように、「夕焼け空」という情報だけで天気を予測することができることから、少ないデータだけでも予測が行なえる可能性があると考えられる。したがって、本論文では様々なデータから必要なデータのみを選択し精度の良い予測を行なう、すなわち、できるだけ小さなシステムで高精度な予測を行なうことを目的とする。小さなシステムを構築することによって、一般のコンピュータでも必要なデータさえ何らかの方法(例えばオンライン)で取得すれば、予測が行なえることが期待できる。

本論文が提案する降雨予測システムは、気象予報における観測データの因果関係を関数近似問題やニューラルネットワークで取り扱うことが必要である⁽⁷⁾と考え、遺伝的前処理を行なうニューラルネットワーク(Neural Network: NN)⁽⁴⁾⁽⁵⁾を用いて降雨予測を行なう。具体的には、過去数日間の局地データを用いて、次の日に雨が降るか否かの局地的降雨予測⁽⁶⁾⁽⁸⁾を行なう。しかしながら実際は、前にも述べたように、予測に用いる多くのデータの中には予測に不必要なデータも数多く存在していることが考えられる。

* 徳島大学大学院工学研究科

〒770-8506 徳島県徳島市南常三島町 2-1

Faculty of Engineering, University of Tokushima

2-1, Minamijosanjima-cho, Tokushima 770-8506, Japan

そこで、データの前処理として、実数値 GA (Real-Coded Genetic Algorithm) を用いることでデータの優先度を決定し、最終的に優先度の高いデータのみを NN の学習に用いることで、予測を行なう。

入力属性の個数が多い場合、出力結果を与える要因の解析は困難である。特徴選択 (入力属性の選択) を行なった研究もいくつか報告されている^{(10) (11)}。文献 (10) では感度解析を用いて、入力属性の選択を行なっているが、入力属性が多い場合の解析は行なわれていない。文献 (11) ではバイナリ GA に基づく方法であり、本論文のように優先度を考慮し解析できない。また、情報基準を用いる中間ユニット数の決定法⁽¹²⁾ も提案されているが、入力属性の選択は行なっていない。入力属性が多い場合、GA による特徴選択と NN による学習と評価が適していると考えられる。したがって、本論文では降雨予測に必要な最小個数の入力属性を決定するためのデータマイニングを実数値 GA により実現する。実数値 GA により必要最小限の因子を決定し、簡単な閾値処理による予測と本論文の方法、さらには全属性を用いる NN 学習の結果を比較している。

本論文では、1998 年の大阪における夏の気象データを用いて計算機シミュレーションを行ない、降雨予測を行なうことで本手法の有効性を定量的に検証する。

2. データ

本論文で扱うデータは、気象庁数値予報課から提供されたものである。データの種類は【平均気温】、【最高気温】、【最低気温】、【平均湿度】、【平均風速】、【最大瞬間風速】、【最大瞬間風速における風向】、【日照時間】、【全天日射量】、【降水量】、【降雪の深さ】、【最深積雪】の計 12 種類のデータを用いる。ここで、このデータの値そのまま用いると、違う種類のデータで同じ値でもその重みが異なってしまう。例えば、最低気温が 20 度と最高気温が 20 度とでは全く違う気候である。これを防ぐために、すべてのデータにおいて正規化を行なう。正規化の範囲は全て 0~1 までの実数値とする。

また、特にデータが大きい場合はその正規化後の値を 1 とする。例えば、台風などで 1 日の降水量が 100 mm を越えるとき、その日の正規化後のデータは 1 とする。また、最大瞬間風速の風向の正規化は、正弦方向と余弦方向とに分ける。したがって、入力属性の個数は 1 日あたり 13 個となる。なお、気象データの中には「測定不能」と表記されるデータが少なくない⁽³⁾。測定不能データとは、データや現象が「0」、つまり全く現象が観測されなかったことを示している。一方、これと似た表現で「0」と表記されているデータがある。「0」は、「観測はされているが、そのデータが 0.001 以上、0.01 未満のもの」を示している。しかしながら、測定不能データと「0」は気象学においてはほぼ同等であるという専門家の意見をもとに、本論文では両データともに「0」としてシミュレーションを行なう。

正規化前のデータを x 、正規化後のデータを x_s とすると、

$$x_s = \frac{x - t_l}{t_u - t_l} \dots \dots \dots (1)$$

の式で、正規化を行なう。ここで、 t_u とは各データの上限、 t_l とは各データの下限とする。各データの上限と下限を表 1 に示す。

3. 降雨予測システム

〈3・1〉ニューラルネットワーク (Neural Network: NN) まず、本論文で用いた NN のモデルを図 1 に示す。ニューラルネットワークの入力は前節で述べた過去 3 日分のデータを入力として与える。したがって、入力信号の個数は合計 39 個となる。出力では、1 を「雨が降る」、0 を「雨が降らない」とする。ここで、「雨が降る」とは、1mm 以上の雨が降ったときのことをいい、「雨が降らない」とはまったく雨が降っていないときのことと定義する。「雨が降る」、「降らない」の判断は、その日の降水量のデータを見て行なう。

〈3・2〉実数値 GA (Real-Coded GA) ニューラルネットワークの学習には不必要なデータも多く存在している。そこで、実数値 GA を用いることによって各入力属性の優先度を決定し、真に予測に必要なデータのみを選択を行なう。本論文で優先度とは、0~1 の範囲であり、対応する入力属性をどれだけ用いるかという割合を表すものである。この優先度が大きいことは、その対応する入力属性の出力に与える影響が大きいと考えることができる。

バイナリーの GA は、0 と 1 の数の並びが個体の遺伝子であるのに対し、実数値 GA は遺伝子の 1 つ 1 つが実数値となり、実数値の並びが 1 つの個体となる。実数値 GA を用いることで属性の使う、使わないのみではなく、その属性がどれだけ降雨に関係しているかの優先度を調べる事が可能である。

実数値 GA の使い方はまず、NN の入力属性と GA の遺伝子とを対応させる。そして、入力属性の優先度がその対

図 1 降雨予測 NN の構造

Fig. 1. A NN structure for rainfall forecast.

表 1 各データの上限と下限

Table 1. Upper and lower bounds of each data.

	Upper bound	Lower bound
Average temperature	32.0	-15.0
Highest temperature	40.0	-10.0
Lowest temperature	30.0	-20.0
Average humidity	100.0	0.0
Mean wind speed	15.0	0.0
Peak gust	35.0	0.0
Sunshine hours	13.0	0.0
Amount of global solar radiation	34.0	0.0
Rainfall level	100.0	0.0
Amount of snowfall	50.0	0.0
Depth of snowfall	90.0	0.0

図 2 1点交叉の例

Fig. 2. An example of one-point crossover.

応する GA の遺伝子とする。優先度は、0 から 1 までの範囲とし、1 に近いものほど優先度は高いものとする。そして、優先度を NN で考慮させるために、入力属性に対して GA で決定した値を乗じる。これは、実数値 GA で決定された値が高いほどその属性が重要であることを示すものである。こうして、NN から得られた結果により適応度を計算し、選択、交叉や突然変異を行ない、次世代の個体の生成を行なう。

ここで、選択の方法はルーレット選択を用いた。また、実数値 GA の交叉は様々存在するが、本論文では、バイナリーの GA の基本的な交叉（1 点交叉）を用いた。また、突然変異はバイナリーの GA では 1 から 0、0 から 1 に変更していたが、実数値 GA では、0 から実数、実数から 0 と変更する。交叉と突然変異の具体的な例を図 2,3 に示す。

〈3・3〉 降雨予測システム 前に述べた NN や実数値 GA を用いた降雨予測システムの手順を以下に示す。

・降雨予測システムの手順

step1 GA における現世代での個体の一つを取り出し、NN の入力属性とその個体の遺伝子とを対応させ、入力属性に対応する遺伝子の値を乗ず。

step2 遺伝子の値を乗じた入力属性によって NN を学習、予測を行なう。

図 3 突然変異の例

Fig. 3. An example of mutation.

step3 NN で予測を行った結果より、適応度を計算する。

この適応度が、その GA の個体の適応度となる。

step4 現世代すべての個体について Step1~3 までの手順を行なう。

step5 すべての個体の適応度から、選択、交叉や突然変異を行ない、次世代の個体の生成を行なう。

step6 決められた世代数まで Step1~5 の作業を繰り返す。最後の世代で、適応度が最大である個体によって決められた優先度による予測を解とする。

また、GA のパラメータは以下に示す。

- ・ 個体数： 20 個
- ・ 世代数： 100 世代
- ・ 遺伝子のビット長： 最大入力属性数
- ・ 交叉の生起確率： 0.35
- ・ 突然変異率：
 - 実数 → 0 0.02
 - 0 → 実数 0.01
- ・ 適応度関数：

$$f = \left\{ \frac{1}{(k_d - x) + \frac{k_d}{k_l}(k_l - y)} \right\}^2 \dots\dots\dots (2)$$

k_d : 予測日数

k_l : 個体の遺伝子のビット長

x : 予測成功数

y : 遺伝子中の 0 の数

ここで、突然変異率と適応度関数は個体の遺伝子中の 0 の個数をできるだけ増やすことによって、予測に必要な入力

属性数を減らす役割も担っている。つまり、入力属性をできるだけ減らすことによって、予測に必要な最小個数の入力属性を決定したいためである。後の計算機シミュレーションで示しているが、不要な入力属性が無い方が認識精度が高くなる。一方、精度が同じであれば（または同等であれば）パラメータの個数が少ない方が優れている⁽¹⁰⁾。また、データマイニング（ルール抽出）にも適している。これが本論文で式(2)を適応度関数に用いた理由である。

4. 計算機シミュレーション

本論文で提案した降雨予測システムの有効性を検証するために、計算機シミュレーションを行なった。

大阪の1998年6月4日から8月2日までの60日間のデータを学習に用い、同地域の1998年8月3日から9月11日までの40日間を予測した。本論文の降雨予測システムの比較のために、NNのみ、すなわちすべての入力属性を用いたときのシミュレーションを行なったところ、40日中24日(60%)予測に成功した。これは、一般に行なわれている天気予報の的中率(約80%)と比べ、予測精度が低いことがわかる。なお、学習データに対しては100%の精度を達成している。

次に、本研究の降雨予測システムのシミュレーションを行なった。5回のシミュレーションを行なったが、その5回とも40日中37日(92.5%)予測に成功した。これは、一般の天気予報と比べ良い結果が得られている。また、予測に失敗した3日は5回とも同じ日で予測に失敗していた。ここで、5回のシミュレーションの1つについて、各世代の適応度の移り変わりを図4に示す。また、この5回のシミュレーションで各入力属性が選ばれた回数を図5、6に示す。ここで、図5、6の横軸は、

- 1: 平均気温 2: 最高気温 3: 最低気温
- 4: 平均湿度 5: 平均風速 6: 最大瞬間風速
- 7: 風向(sin) 8: 風向(cos) 9: 日照時間
- 10: 全天日射量 11: 降水量 12: 降雪の深さ
- 13: 最深積雪

とする。図4を見てもわかるように、徐々に適応度の高い個体が残っていることがわかる。また、図5から、優先度が0より上、すなわち入力属性がよく使われたものを取り上げると、

3日前: 最低気温, 最大瞬間風速の風向(正弦方向と余弦方向)

であった。

また、図6よりGAにより得られた優先度が0.5以上、すなわち入力属性が半分以上使われた回数が多いものを取り上げると、

3日前: 最大瞬間風速の風向(正弦方向)

2日前: 日照時間

1日前: 平均気温, 日照時間

図4 適応度の移り変わり

Fig. 4. Change of fitness value.

図5 (優先度)>0で選ばれた各入力属性の回数

Fig. 5. The number of survived attributes with priority greater than 0.

図6 (優先度)≥0.5で選ばれた各入力属性の回数

Fig. 6. The number of survived attributes with priority greater than 0.5

であった。以上のこれら選ばれた入力属性は、降雨を予測するためには必要な入力属性であると考えられる。そこで、これら6つの入力属性のみを用いて予測を行なった。優先度は、使われたうちの優先度の平均値とする。このときのNNの構造を図7に示す。その結果、40日中37日(92.5%)

図 7 最終的な NN の構造

Fig. 7. Structure of final NN.

表 2 各属性の平均値

Table 2. Average values of each attribute.

Attribute	Not raining	Raining
Lowest temperature (3 days ago)	23.8(°C)	22.8(°C)
Sunshine hours (2 days ago)	5.83(h)	4.08(h)
Sunshine hours (a day ago)	5.45(h)	4.89(h)
Average temperature (a day ago)	27.0(°C)	25.8(°C)

表 3 各属性の閾値

Table 3. Threshold values of each attribute.

Attribute	threshold value
Lowest temperature (3 days ago)	23.3
Sunshine hours (2 days ago)	4.955
Sunshine hours (a day ago)	5.17
Average temperature (a day ago)	26.4
Wind direction of peak gust (3 days ago)	north-northwest and south-southeast

予測に成功した。本論文での GA は入力属性の感度や属性間の相関などを考慮せず、認識精度を改善するための属性の組み合わせを選択している。したがって優先度の値が大きい属性ほど精度改善に貢献したと考えられる。参考のために、過去 5 日間と 7 日間のデータを用いて同様の計算機シミュレーションを行なった。その結果、5 日間では残った入力属性は 11 個で精度は 92.5%，7 日間では残った入力属性は 19 個で 90.0% となった。したがって、一番効率の良いのは 3 日間のデータを用いて行なったときである。

ところで、選ばれた入力属性について、実際に雨が降っている日とそうでない日のデータの平均値を調べてみると、表 2 の通りであった。また、最大瞬間風速の風向き（3 日前）について、雨が降るときは西南西の風、雨が降らないときは東北東の風がよく吹くことがわかった。この平均値の中央値を閾値とする降雨予測を行なった。閾値の値は、表 3 に示す。閾値処理の方法は、5 つの入力属性それぞれ単独に閾値処理を行ない、5 つの多数決降雨予測を行なうも

のである。その結果、40 日中 22 日（55%）成功した。したがって、本研究の降雨予測システムの方が有効であることが言える。この場合、予測精度が 55% 程度であったことは、データマイニングの観点からも重要な意味を持つ。つまり、1 属性の大小関係だけでは高精度の降雨予測は困難であることを意味している。予測精度を上げるためには複数属性の組み合わせからなる予測方式が必要であり、そのための予測システムとして NN が効果的であることが結果として示されたことになる。なお、1 属性からなる閾値処理ルールは人間が最も理解しやすい形式ではあるが、複数属性の組み合わせが高精度を達成するために必要であることはすでに報告されている⁽⁷⁾。

最後に、選ばれたデータについて考える。低気温のデータを見ると、雨が降っている日の 3 日前は雨が降っていない 3 日前と比べ、平均して約 1℃ 気温が低い。また、2 日前の日照時間も雨が降った日の 1, 2 日前は日照時間が約 2 時間低く、1 日前の日照時間では約 0.5 時間低い。1 日前の平均気温も同じく、雨が降っていない日と比べ約 1℃ 低い。最低気温、日照時間そして平均気温は、雲の量に大きく影響される。雨は雲の量に大きく影響するため、最低気温、日照時間や平均気温が予測に必要であると考えられる。最大瞬間風速の風向きが必要な理由は、風向きによって現在のあたりに低気圧または前線があるのかという情報が得られるためであると考えられる。参考のため、他の季節でも同様のシミュレーションを行い、結果的に同じ 6 属性が選択されることを確認した。

5. 結 言

本論文では実数値 GA と NN を用いた降雨予測システムを提案した。実数値 GA により降雨予測に必要な入力属性を選択し、選択された属性のみで降雨予測を行った。その結果、92.5% 予測成功した。これは、NN のみで行なった結

果(60.0%)や現在行なわれている天気予報の精度(約80%)よりも良い結果が得られた。したがって、本論文の降雨予測システムの有効性を検証することができた。

本論文では、大阪の夏の季節のみしかシミュレーションを行っていないため、今後の予定として、すべての季節において予測を行なうことが挙げられる。そして、大阪以外の地域の予測も行ないたいと考えている。また、短時間後の予測や長期予測などを行なう予定である。

(平成13年10月2日受付, 平成14年11月15日再受付)

文 献

- (1) 饒村 曜: イラストでわかる天気のはくみ, 新星出版社 (1999)
- (2) 安居院猛, はか: ジェネティックアルゴリズム, 昭晃堂 (1993)
- (3) Y. Mitsukura, M. Fukumi, and N. Akamatsu: "Fog Occurrence Forecast by Using LVQ with a Genetical Preprocessing", *Trans IEE Japan*, Vol. 120-C, No. 12, pp. 2055-2061 (2000-12)(in Japanese)
満倉靖恵・福見 稔・赤松則男:「遺伝的処理を行う LVQ を用いた霧発生予測法」, 電学論 C 120, 12, pp.2055-2061(2000-12)
- (4) 中野 肇, はか: 入門と実習ニューロコンピュータ, 技術評論社 (1989)
- (5) 馬場則夫, はか: ニューラルネットの基礎と応用, 共立出版 (1994)
- (6) H. Nakayama, N. Abe, and S. Hattori: "The meteorological prediction ability on the hierarchical neural network", *Technical Report of IEICE*, NC92-10, pp.67-74 (1992)(in Japanese)
中山 仁・安部倫之・服部進実:「天気予測に見る階層型ニューラルネットワークの予測能力」, 信学技法 NC92-10, pp.67-74(1992)
- (7) M. Fukumi and Y. Mitsukura: "Knowledge Insertion and Rule Extraction in Neural Networks", *S2PATJ*, Vol. 12, pp. 13-18 (2001-12)(in Japanese)
福見 稔・満倉靖恵:「ニューラルネットワークへの知識の埋込とルール抽出」, 高速信号処理応用学会論誌(電子技術), 12, pp.13-18(2001)
- (8) K. Ochiai, H. Suzuki, K. Shinozawa, M. Fujii, and N. Sonohara: "Rainfall and Snowfall Prediction from Images of Weather Radar by using Nwural Network Model", *Technical Report of IEICE*, NC95-15, pp.111-118 (1995-5)(in Japanese)
落合慶広, はか:「ニューラルネットワークを用いた気象レーダ画像による降雨・降雪予測」, 信学技報, NC95-15, pp.111-118(1995-5)
- (9) H. Akaike: "A New Look at the Statistical Model Identification", *IEEE T. on Automatic Control*, Vol. AC-19, No.6, pp.716-723(1974)
- (10) H. Takenaga, S. Abe, M. Takatoo, M. Kayama, T. Kitamura, and Y. Okuyama: "Input Layer Optimization of Neural Networks by Sensitivity Analysis and its Application to Recognition of Numerals", *T. IEE Japan*, Vol.111-D, No.1, pp.36-44(1991-1)(in Japanese)
武長 寛, はか:「感度解析を用いたニューラルネットワークの入力層の最適化とその数字認識への応用」, 電学論 D 111, 1, pp.36-44(1991-1)
- (11) M. Fukumi, N. Akamatsu: "A Genetic Algorithm with Deterministic Mutation Based on Neural Network Learning", *IEICE*, Vol.J80-D-II, pp.2800-2807(1997)(in Japanese)
福見 稔・赤松則男:「ニューラルネットワークの学習に基づいた決定論的変異を有する遺伝的アルゴリズム」, 信学論, J80-D-II, pp.2800-2807(1997)
- (12) 塚本浩弘: 明日の天気分かる本, 地球丸 (1998)
- (13) 栗田多喜夫:「情報量基準による3層ニューラルネットワークの隠れ層のユニット数の決定法」, 信学論, J73-D-II, 11, pp.1872-1878(1990-11)

伊藤 征嗣 (学生員) 2001年3月徳島大・工・知能情報工卒, 現在, 同大学大学院博士前期過程在学中, 信号処理, 生物型情報処理などの研究に従事。

満倉 靖恵 (正員) 2001年9月徳島大学大学院工学研究科博士後期課程期間短縮修了, 1999年4月徳島大学工学部知能情報工学科助手, 現在に至る。信号処理, 生物型情報処理などの研究に従事, 博士(工学)

福見 稔 (正員) 1987年3月徳島大学大学院工学研究科修了, 同年4月同大情報工学科助手, 1995年同知能情報工学科講師, 1996年同助教授, 現在に至る。信号処理, 生物型情報処理などの研究に従事, 1995年計測自動制御学会論文賞, 京大工博。

赤松 則男 (非会員) 昭和41年徳島大・工・電気卒, 1971年京大大学院博士課程修了, 同年徳島大・工・助手, 1975年同助教授, 1987年同教授, 京大工博, 生体情報工学, 非線形回路, 音声の認識と合成, 電子ペンの研究に従事。