

ニューラルネットワークによる説明変数をもつ時系列データの予測における学習データの選定方法

4K-2

下平 丕作士

日本メックス(株) 研究開発部

1. まえがき

多層前向きニューラルネットワークにより説明変数を持つ時系列データの予測を行う際の、学習データの選定方法について検討した。空調の予熱・予冷時間に関するデータを用いて数値実験を行った。

2. 予測方法と学習データの選定方法

応答変数 y_t に対して1つの説明変数 x_t がある場合について説明する。応答変数については y_{t-1} までの、説明変数については x_t までの値が計測されているものとし、 y_t を予測する場合を扱う。式(1)のように、説明変数と応答変数についてそれぞれ d 個及び $(d-1)$ 個の予測用データグループを入力として、予測値として Y_t を出力する。このような予測値を出力するために、過去のデータを学習してネットワークの重みを求める。学習時には、式(2)の左辺のような学習用データグループを入力として、 $Y_{t'}$ を出力する。誤差逆伝播法により、 n 個の学習データグループを用いて、出力値 $Y_{t'}$ と計測値 $y_{t'}$ の差が最小になるように、重みの値を求める。

従来からよく用いられている移動窓データ学習法(MWDL法: Moving Window Data Learning Method)では、時点 t の直前に一定の大きさの窓を設定し、その範囲のデータを式(2)のようにグループ化して全て学習する。ここでは、 n と d によって移動窓の大きさを定義する。

類似データ選定学習法(SDSL法: Similar Data Selective Learning Method)では、予測用データグループに類似したデータグループを、学習データ選定範囲内から類似度が高い順に n 個選定して学習する。類似度の尺度として、重み付き距離を用い、距離が小さいほど類似度が高いと考える。筆者¹⁾は、距離の重みとして、式(3)、(4)のように時系列データの相関係数の巾乗の関数を用いる方法(CSDS法: Correlation Based Similar Data Selection Method)を提案した。式(3)は、予測用と学習用データグループの間の重み付きマンハッタン距離、

$$(x_{t-d+1}, \dots, x_{t-1}, x_t, y_{t-d+1}, \dots, y_{t-2}, y_{t-1}) \rightarrow Y_t \quad (1)$$

$$(x_{t'-d+1}, \dots, x_{t'-1}, x_{t'}, y_{t'-d+1}, \dots, y_{t'-2}, y_{t'-1}) \rightarrow Y_{t'} \quad (2)$$

$$D = \sum_{i=0}^{d-1} |\rho_{x_i}|^m |x_{t-i}^s - x_{t'-i}^s| + \sum_{i=1}^{d-1} |\rho_{y_i}|^m |y_{t-i}^s - y_{t'-i}^s| \quad (3)$$

$$D = \left\{ \sum_{i=0}^{d-1} |\rho_{x_i}|^m (x_{t-i}^s - x_{t'-i}^s)^2 \right\}^{\frac{1}{2}} + \left\{ \sum_{i=1}^{d-1} |\rho_{y_i}|^m (y_{t-i}^s - y_{t'-i}^s)^2 \right\}^{\frac{1}{2}} \quad (4)$$

図1 本文中の式

式(4)は重み付きユークリッド距離である。 ρ_{x_i} と ρ_{y_i} は、それぞれ y_t と x_{t-1} 及び y_t と y_{t-1} の間の相関係数であり、学習データ選定範囲内のデータを用いて計算する。 y_t が未知なので、これを除いたデータの組によって計算する。 ρ_{y_i} と ρ_{x_i} は、 y_t がそれぞれ y_{t-1} と x_{t-1} に依存する度合いを表す。式(3)と(4)では、これらの値が大きいほど、距離の計算においてその座標値を大きく扱う。これにより、予測すべき値により強い影響を持つデータを重視した距離が計算される。

3. 数値実験

3層前向き型ニューラルネットワークを用いた。誤差逆伝播法の収束条件は、出力ノードの誤差の2乗和の $1/(2n)$ が、予熱時間については0.001、予冷時間については0.0005以下になったときとした。各パラメータの値を少しずつ変えてシミュレーションを行い、予測精度を調べた。

予熱時間については山形市内のAビル、予冷時間については東京都内のBビルで1時間毎に計測された特定の部屋の温度、湿度等のデータを用いた。これらのデータを用いて、直線補間により目標温度等になるのに要する時間を求めて、近似的な予熱・予冷時間として用いた。予測日を t とし、2. で述べたようにして予測値を求め、計測値と比較した。空調装置は休日には運転されないもので、休日は予測の対象から除いた。SDSL法では、学習デ

A Method of Selecting Learning Data in the Prediction of Time Series with Explanatory Variables Using Neural Networks

Nihon MECCS Co., Ltd.

1-22-10 Nishi-Shinbashi, Minato-Ku, Tokyo 105, Japan

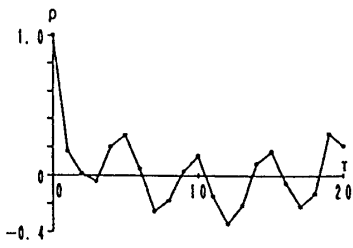


図2 予熱時間の自己相関係数

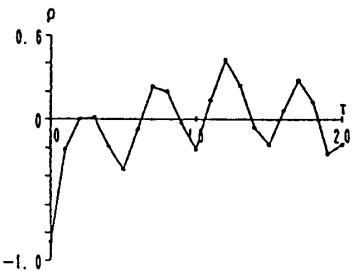


図3 午前8時の室温の相関係数

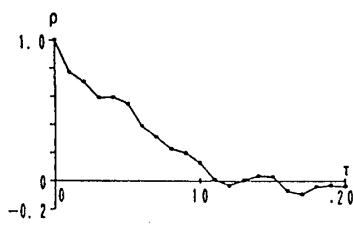


図5 予冷時間の自己相関係数

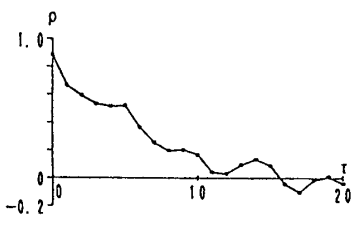
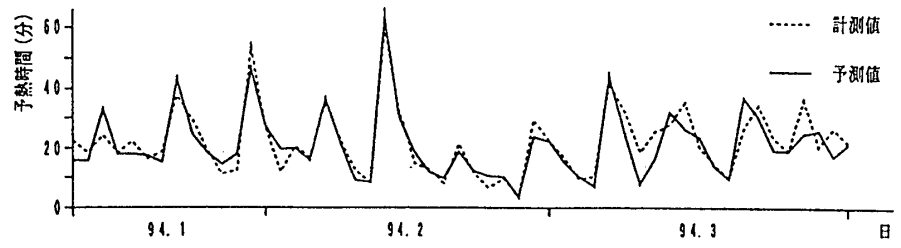
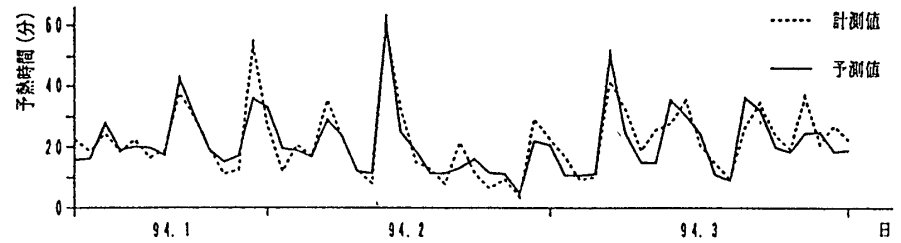


図6 午前7時の室内不快指数の相関係数

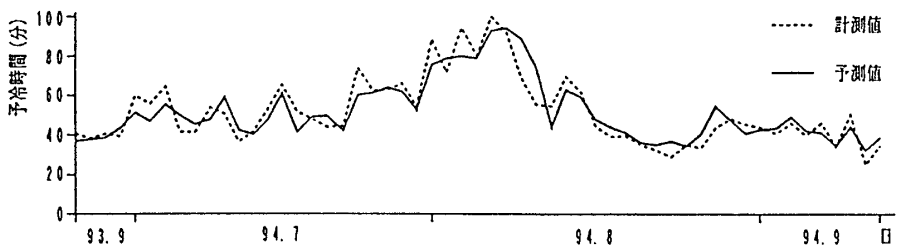


(a) SDSL法 (7-11-1, $d=2$, $n=20$, ユークリッド距離, $m=2$)

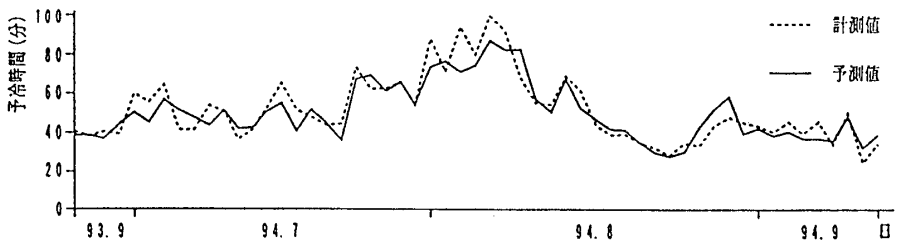


(b) MWDL法 (7-8-1, $d=2$, $n=22$)

図4 予熱時間の計測値の時系列データとその予測値の比較



(a) SDSL法 (23-3-1, $d=6$, $n=22$, マンハッタン距離, $m=6$)



(b) MWDL法 (23-1-1, $d=6$, $n=19$)

図7 予冷時間の計測値の時系列データとその予測値の比較

一夕選定範囲を40日とし、予測日に伴って移動させた。

予熱時間の予測には1993年から1994年にかけての冬期の76日分のデータを用いた。目標温度を20℃とした。説明変数として、当日の午前8時の室内の温度と湿度、及び前日の午前12時の室温を選定した。前日の午前12時の室温は空調装置が運転されていたかどうかとともに、建物の冷え具合を考慮するために用いた。図2に自己相関係数、図3に午前8時の室内の温度との相関係数を示す。図4に計測値と予測精度が最もよい場合の予測値を示す。

予冷時間の予測には1993年及び1994年の夏期の82日分のデータを用いた。応答変数を室内の不快指数とし、目標値を75とした。説明変数として、当日の午前7時の室

内と外気の不快指数、及び前日の午前12時の不快指数を選定した。図5に自己相関係数、図6に午前7時の室内の不快指数との相関係数を示す。図7に計測値と予測精度が最もよい場合の予測値を示す。

4. むすび

学習データ選定方法による予測精度は、時系列データの性質に依存することが分かった。変わりやすい性質の予熱時間の場合には、SDSL法の方がMWDL法よりも予測精度がかなりよい。滑らかに変化する性質の予冷時間の場合には、MWDL法の方がSDSL法よりも予測精度がややよい。