

配水量予測への学習手法の適用

玉田 隆史* 丸山 稔* 阿部 茂* 前田 和男**

*三菱電機株式会社 中央研究所

**三菱電機株式会社 制御製作所

概要

与えられた例題（学習データ）からなるデータベース上にデータ間の距離を定義し、入力を与えられた時に、入力データと距離の近い例題を用いて入出力の関係を求めることで推論 (Memory-Based Reasoning) や学習 (Memory-Based Learning: MBL) を行なう方式が提案されている。本論文では、MBL と重回帰モデルを併用した予測モデル (MBL モデルと呼ぶ) を提案し、浄水場の日配水量予測への適用を行なった。MBL モデルは、入力データが与えられた時に、例題をもとに入力データの近傍を定義し、近傍内に例題が存在する場合には MBL で予測を行ない、存在しない場合にはグローバルモデルで予測を行なう方式である。

予測実験によれば、MBL モデルは重回帰モデルに比べ予測精度が 5% 程度向上し、特に、近傍内にデータが存在する場合には、10-30% 向上した。

キーワード

日配水量予測 Memory-Based Learning 重回帰分析 学習手法 ニューラルネット

1. はじめに

浄水の安定供給と送配水系統及び浄水場の効率的運用を図るには、配水量の需要予測を精度良く行なう必要がある。需要予測には、年・月単位の長期水需要予測と、日・時間単位の短期水需要予測があり、長期水需要予測は需給関係の変化にともなう施設の拡張計画に、短期水需要予測は最適な配水コントロールを計画する際に利用される。本論文では、短期水需要予測である日配水量予測について述べる。

従来、日配水量予測には、重回帰モデル等の統計的モデルを用いた手法が主に提案されてきた [1,2]。しかし、与えられた有限個のデータから、全データに適合する適切なモデルを決定するのは一般に困難であることが多い [3]。近年、データからの学習に基づいてモデルを決定する学習手法の予測問題への適用が電力需要予測の分野を中心に盛んに試みられている。特に、ニューラルネットは非線形な関係を近似する能力が優れているとして適用例が多い [4]。一方我々は、データの近接関係に基づいて推論 [5] や学習 [6] を行なう Memory-Based Learning (MBL) の日配水量予測への適用を検討している [7]。MBL は入力を与えられた時に、入力データと距離の近いデータを検索し、検索したデータを用いて入出力の関係を求めて出力を行なう学習手法であり、ロボティクスの分野等で研究が行なわれている [8]。ニューラルネットが与えられた全データに適合するようなモデル (グローバルモデル) の学習を行なうのに対し、MBL は一部のデータに適合するようなモデル (ローカルモデル) の学習を行なうという特徴がある。

本論文において、MBL の日配水量予測への適用可能性を検討し、MBL と重回帰モデルを併用する方式を提案する。本方式によれば、重回帰モデルのみによる予測と比較して精度が 5% 程度向上した。

2. 日配水量の特性

図 1 は某浄水場の日配水量のトレンドグラフである。この図から分かるように日配水量は変動要因の影響を受け日々変動していく。従って、予測モデルを構築する際に、日配水量の変動に影響を与えている変動要因を明らかにし、説明変数 (予測モデルにおける入力変数) として何を選択すべきかを決定する必要がある。AIC 評価基準により、最高気温、天候、同天候の最近日の配水量の組を日配水量の説明変数として選択した [9]。

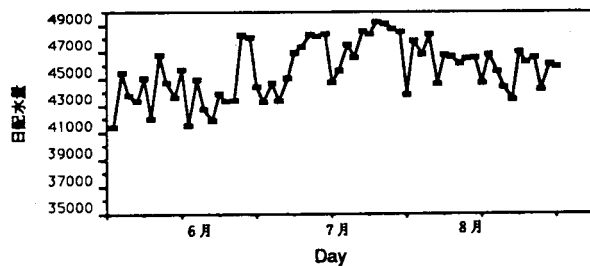
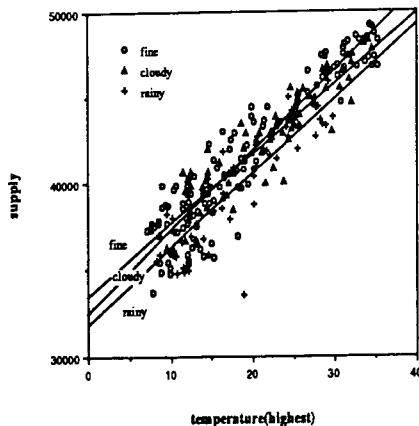
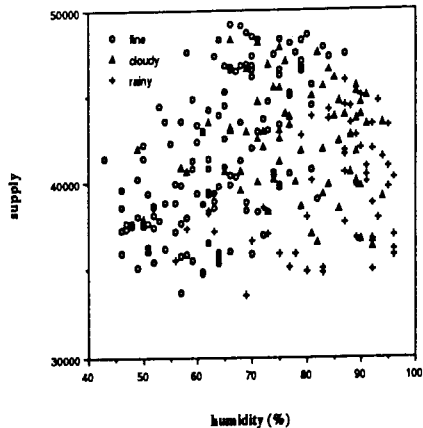


図 1 日配水量のトレンドグラフ



(a) 最高気温と日配水量



(b) 湿度と日配水量

図2 説明変数候補(温度、天候、湿度)と日配水量の関係

図2は3つの変動要因候補(最高気温、天候、湿度)と日配水量の関係をプロットしたものである。図から、

- ・最高気温と日配水量の間には線形性が見られる。
- ・日配水量は天候により変動する。(洗濯为天候の影響を受けることによる)
- ・湿度と日配水量の間には相関性がみられない。

ことがいえ、最高気温や天候は説明変数として妥当であると考えられる。

3. MBLを用いた日配水量予測

3.1 Memory-Based Learning

MBLは過去の例題(入力データと出力データの組)からなるデータベース上に例題間の距離をデータの分布に応じて定義し、与えられた問題と距離の近い例題を検索し、検索した例題をもとに出力を決定する。従って、距離が近い場合には出力データの値も近くなるように距離を定義していかなければならない。一般に、例題を追加することにより、入力データと距離の近い例題が存在する確率も高くなるので、MBLでは例題数が多くなるほど、出力値は正しい出力へと近づいていく。一方、データ数や次元数の増加は検索コストの増大をともない、検索効率が悪化する。MBLではデータベースへの範囲検索が頻繁に行なわれるので、多次元データの効率的な管理・検索が可能なMD木[10]等によりデータを管理する。

従って、MBLにおいては、新たに例題を記憶し、検索が高速となる様に例題を構造化して記憶することが学習に相当する。図3に、入力変数が2次元である場合のMemory-Based Learningのイメージを示す。入力データ(O)に対する出力値 y は斜線内の領域に含まれる例題($X1, X2$)を用いて $y=f(X1, X2)$ として決定する。

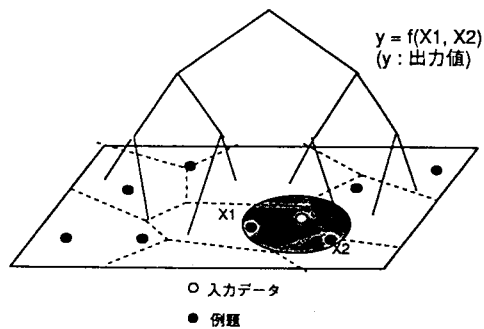


図3 Memory-Based Learning

3.2 MBLを用いた予測モデル

MBLは入力データと十分距離の近いデータが存在する場合には、精度の高い予測が期待できる。しかし、距離の近いデータが存在しない場合、MBLでは十分な精度の予測が困難となる。

そこで我々は、入力データが与えられた時にMBLのモデリング能力の高い領域(近傍)を定義し、近傍内にデータが存在する場合にはMBLにより予測を行ない、近傍内にデータが存在しない場合にはグローバルモデルにより予測を行なう予測モデル(MBLモデル)を考えたい。次節でまず、近傍の定義方式を述べる。

3.3 近傍の定義

一般に、説明変数の目的変数への説明能力（感度）は説明変数ごとに異なる。従って、近傍を定義するとき、各説明変数の感度に応じて重み付けを行なう必要がある。また、特異データの存在も考慮しなければならない。以上の点をふまえて、入力データ p の近傍 $U(p)$ として、 $U(p)$ 内のほとんどのデータに対応する出力（日配水量） y_λ と予測日の配水量 y との誤差 $\frac{|y_\lambda - y|}{y}$ が一定の範囲 E 内となるように以下の手順で定義する。

今、 X_1, X_2, \dots, X_m を予測に使用する説明変数、 Y を目的変数、 $p = (x_1, x_2, \dots, x_m), x_i \in X_i$ を入力データ、 $\bar{x}_j = (x_1^j, x_2^j, \dots, x_m^j), x_i^j \in X_i^j$ を j 日目の説明変数の値（例題データ）、 \bar{y}_j を j 日目の日配水量、 $d_j = (\bar{x}_j, \bar{y}_j)$ とする。

(1) 近傍 $U(p)$ の決定

予測モデルに要求する誤差を E とした時に、 E の値に応じて説明変数の重み付けを行ない、入力データの近傍を求める。

$j(1 \leq j \leq n)$ 日目のデータ d_j に対し、各説明変数 X_i ごとに

$$|j-k| < n, \frac{|y_j - y_k|}{y_j} < E \Rightarrow |x_i^j - x_i^k| < \bar{\beta}_i^j \quad (2)$$

を満たすような $\bar{\beta}_i^j$ の最小値 β_i^j を求める。ここで(2)式の左辺を満足する様なデータが存在しない場合、 $\beta_i^j = \infty$ とする。この時、 β_i^j は j 日目における変数 X_i の感度を表している。ここで、 $\hat{\beta}_i = \min_{(1 \leq j \leq n)} \beta_i^j$ を予測日における変数 X_i の感度と考え、 p の近傍 $U(p)$ を

$$U(p) = (x_1 - \hat{\beta}_1, x_1 + \hat{\beta}_1) \times \dots \times (x_m - \hat{\beta}_m, x_m + \hat{\beta}_m)$$

と定義する。但し、 $\beta_i^j = \infty (1 \leq j \leq n)$ の時は、 $U(p) = \phi$ とする。図4に、説明変数が2次元の場合の近傍定義のイメージを示す。

(2) 特異データの除去

$\overline{UD(p)}$ を例題データが $U(p)$ に含まれる予測日から n 日以内の実績データからなる集合とする。この時、 $\overline{UD(p)}$ の元 d_j で、 $\beta_i^j = \infty (1 \leq i \leq m)$ となるものは特異データとみなし、 $\overline{UD(p)}$ から特異データを除いたものを $UD(p)$ として $UD(p)$ を用いて出力 y を求める。即ち、 $y = f(UD(p))$ となる。ここで、関数 f としては平均や重回帰式を使用する。

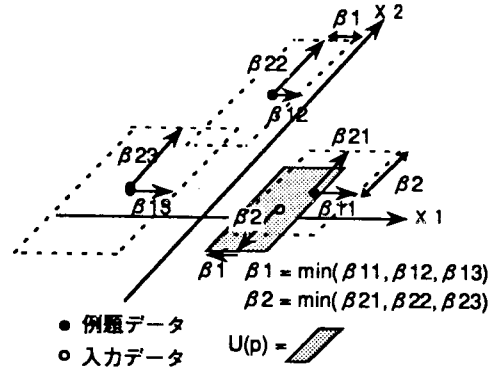


図4 近傍の定義

3.4 MBL モデルの予測アルゴリズム

E を予測モデルに要求する精度、 n を近傍を定義する時に参照する日数、 D は重回帰モデルでのデータ参照数とする。近傍データが存在しない場合のグローバルモデルとしては、2.より最高気温と日配水量の関係に線形性がみられるため重回帰モデルを用いる。この時 MBL モデルは、以下のアルゴリズムで予測日配水量 y を求める。

- A1. 与えられた n, E に対し、3.3(1)の方法により p の近傍 $U(p)$ を決定する。
- A2. $U(p)$ に含まれる近傍データ $\overline{UD(P)}$ を求める。 $\overline{UD(P)}$ から特異データを除き $UD(p)$ を求める。
- A3. $UD(p) \neq \phi$ の時、 $UD(P)$ 内のデータの日配水量の平均を予測値とする。
- A4. $UD(p) = \phi$ の時、過去 D 日のデータを用いて重回帰モデルにより予測値を求める。

4. 予測実験

4.1 実験条件

MBL モデル (MBL + 重回帰) と重回帰モデルとの比較を行なった。実験に使用したデータは某浄水場のある年の4月から12月までの平日の日配水量であり、そのうち6月から12月までの平日 (139データ) を予測対象日とし、両モデルで予測を行ない比較した。

説明変数としては、2章で述べた点と、日配水量と説明変数の関係が時変系 (季節により関係が変化) である点を考慮して最高気温、当日の天候 (晴れ、曇り、雨)、前予測日の天候 (晴れ、曇り、雨)、同天候の最近日の配水量、日幅を使用し、 $E = 1.0, 1.5 (\%)$, $D = 40, n = 10, 15, 20$ に対し各々実験を行なった。

表1 MBL適用日における予測誤差比較

(a) E=1.0%の時

	MBL(%)	重回帰(%)	DD
n=10	1.45	2.31	14
n=15	1.03	1.51	11
n=20	1.31	1.41	14

(b) E=2.0%の時

	MBL(%)	重回帰(%)	DD
n=10	1.34	1.53	33
n=15	1.24	1.46	40
n=20	1.45	1.45	45

表2 予測結果(誤差)

	MBLモデル	
	E=1.0%	E=2.0%
n=10	1.66	1.69
n=15	1.70	1.68
n=20	1.73	1.74

重回帰モデル(%)	D
1.74	40

表3 要求精度(E)とMBLの関係

要求精度 E (%)	誤差平均 (%)	DD
2.0	1.24	40
1.5	1.13	25
1.0	1.03	11

4.2 結果と考察

表1にMBLモデルにおいてMBLを用いて予測を行なった日数(DD)と、その時の平均予測誤差を同じ日に重回帰モデルを適用した結果と比較したものを、表2に各モデルの全予測日における平均予測誤差を示す。

表1より、MBLにより予測を行なった日では、予測精度はおおむね目標値(E)をみたくており、予測精度が10-30%向上している。しかし、Eの値が小さい場合にはMBLの適用されるケースが少なくなり、nに依存して精度が変動する。表2は、MBLモデルと重回帰モデルによる全データに対する予測結果である。全てのケースにおいてMBLモデルの方が重回帰モデルに比べ精度が良くなっている。

MBLと要求精度Eの関係を調べるために、n=15と固定し、E=1.0, 1.5, 2.0%におけるMBLの平均予測誤差を求めた(表3)。表3より、Eの値を小さくすることにより、予測精度は良くなっていくが、MBLにより予測を行なう日数は減少する。これは、Eの値を小さくすることにより、近傍条件が厳しくなるためと考えられる。また、各ケースにおいて予測精度はほぼ要求精度を満たしており、近傍の定義の有用性が示された。以上のことから、MBLの精度はEの値に比例して変動していくので、要求する予測精度や併用するグローバルモデルの精度をもとEの値を設定する必要がある。

5. おわりに

本稿では、MBLの日配水量予測への適用を検討し、MBLとグローバルモデル(重回帰分析)を併用する予測モデルを提案した。MBLモデルを日配水量予測に適用した結果、重回帰モデルのみによる予測に比べ精度が向上し、MBLモデルの有用性が示された。

今後は、他の浄水場へのMBLモデルの適用を行なうとともに、重回帰モデル以外のグローバルモデル(ニューラルネット等)とMBLを併用したモデルの検討、他の予測問題への適用を行なう予定である。

参考文献

- [1] 齊藤、小林、"小特集 3. 需要予測と水量配分計画"、電気学会誌, Vol.101, No.3, pp.14-17(1981)
- [2] 神野、河村、上田、"福岡市の浄水場における日配水量の変動特性と予測について"、九大工学集報, Vol.59, NO4, pp.495-502(昭61)
- [3] J. Doyne Farmer and John J. Sidorowich, "Prediction Chaotic Time Series", Phys. Rev. Lett., Vol. 59, No. 8, pp.845-848(1987)
- [4] D.C. Park, "Electric Load Forecasting Using An Artificial Neural Network", IEEE trans. Power System, Vol.6, No.2, pp.442-449(1991)
- [5] Craig Stanfill and David Walts, "Toward memory-based reasoning", CACM, Vol.29, No.12, pp.1213-1228(1986)
- [6] S.M. Omohundro, "Geometric Learning Algorithms", Physica D, Vol. 42, pp.307-321(1990)
- [7] 玉田、丸山、中村、阿部、"Memory-Based Learning による配水量予測"、平成4年電気学会全国大会 1549
- [8] 有本、"ロボットと学習"、人工知能学会誌, Vol.5, No.6, pp.748-756(1990)
- [9] 丸山、玉田、他、"Memory-Based Learning を用いた配水量予測"、電学研資、IP-92-4(1992)
- [10] 中村、他、"多次元データの平衡木による管理—MD木の提案—、信学論、J71-D,9, pp.1745-1752(1988)