

3. アメダス、ランドサットデータを用いた気象要素の推定手法についての研究

1. 背景と目的

メッシュ気候値の作成のためには、重回帰分析やニューラルネットワークが、測定地点間のデータを補間し推定を行う手法として用いられている。このような手法では、観測地点のないところの、気温などを地形因子から推定するため、地形因子を独立変数とした重回帰分析や、非線形重回帰手法としてニューラルネットワークが使用されてきた。従来の研究では、ニューラルネットワークの利点として、線形重回帰の場合問題となる多重共線性の問題がないことや、非線形な関係にあるデータの回帰に適しているなどがあげられている。しかし、両手法の厳密な比較には、cross validation などの手法で詳細に検討する必要があるが、そのような例は少ない。

そこで、本研究では、Leave-one-out cross validation により、重回帰分析とニューラルネットワークの精度比較を行うことを目的とする。

2. 研究の方法

Leave-one-out cross validation 法とは、全サンプルから、1つのサンプルを除き、残りのサンプルでモデルを作製したときに、あらかじめ抜いていたサンプルを正確に予測することができるかを検定する方法である。本研究では、ニューラルネットワークと重回帰分析の誤差は、この Leave-one-out cross validation 法を用いて、全箇所すべてのデータを使用して推定式の作成を行い、1ヶ所のデータを検証用にとっておき、残りのデータを使用して両者の推定式を計算した後、検証用データの推定値を求めた。この作業をすべての箇所について行った。

推定する気候値のデータは、図1のように北緯34°～36°40′，東経138°～141°内の、主に関東地方にある91ヶ所の観測所における、アメダス気温データの平年値を用いた。

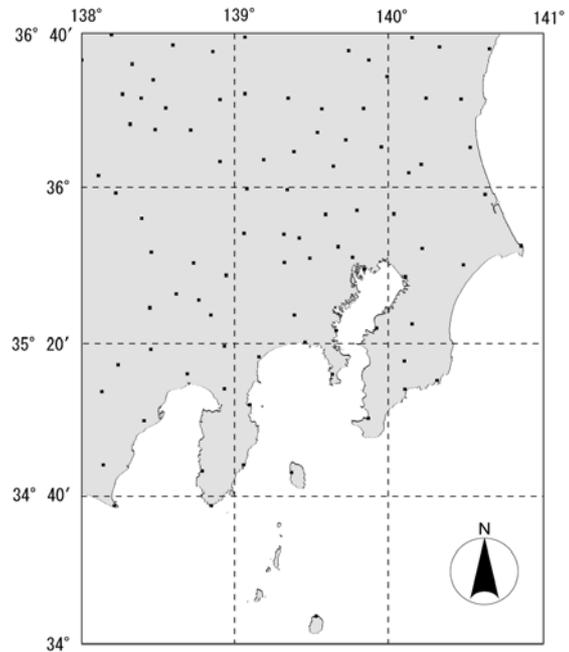


図1 観測所所在地

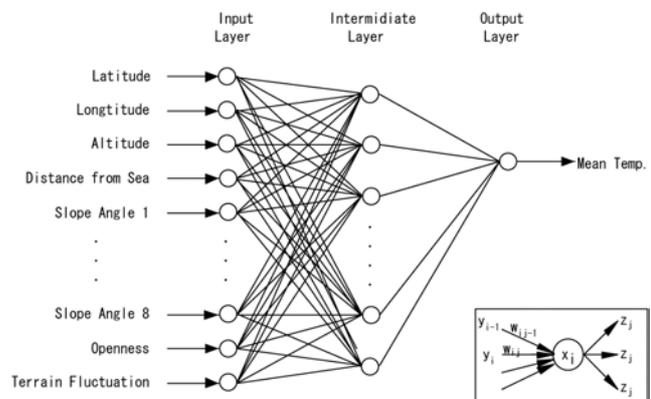


図2 ニューラルネットワークの模式図

地形因子の算出には、国土数値情報 250m メッシュ標高データ（国土地理院）を使用した。地形因子は、既往の研究などを参考に、比較的気候値へ及ぼす影響が大きいと考えられる、緯度、経度、標高、海岸距離、傾斜度、露出度、有効起伏量を用いた。

推定に使用したニューラルネットは、地形因子を入力層、気候値を出力層とし、中間層は一層からなるものとした。入力層、出力層の数は、使用した地形因子、気候値の数に応じて、それぞれ 14 ユニット、1 ユニットとした(図 2)。中間層のユニット数は、入出力層の倍、同数、1/2、1/4 などの場合から推定精度を試行錯誤的に比較し、14 ユニットとした。

入力にあたりデータは 0~1 の範囲にスケールした。学習には、バックプロパゲーションアルゴリズムを用いた。バックプロパゲーションアルゴリズムでは、各ユニットの入力値、出力値との関係は次のように表わされる（図 2 参照）。

$$x_j = \sum_i w_{ij} \times y_i \quad (1)$$

$$z_j = \frac{1}{1 + e^{-x_j}} \quad (2)$$

ここで、 y_i はユニット j への入力値、 z_j は j からの出力値、 x_j は内部状態、 w_{ij} はユニット間の結合の重みを表わす。

推定値と教師値（測定値）との誤差 E は、次式のように定義されるが、学習は、誤差 E を最小化するように、 w_{ij} を最適化することにより行われる。

$$E = \sum_k (d_k - g_k)^2 \quad (3)$$

ここで、 g_k は出力層の k 番目のユニットの出力値（推定値）、 d_k は教師値（測定値）である。

最終的に求められた w_{ij} と、式(1)、(2)式により推定式が表わされる。

プログラムは、Visual Basic(Microsoft)により作成した。重回帰分析のソフトウェアには、R を用いた。従って、推定式および推定値は、ニューラルネットワークと重回帰分析ともに、91 箇所の場合について算出し、誤差の評価を行った。

Leave-one-out cross validation のためのプログラムについては、ニューラルネットは、Visual Basic(Microsoft)により作成し、重回帰分析は R により作成した。誤差は、Root Mean-Square Error (RMSE) により評価した。図 3 にニューラルネット、重回帰分析を使用して作成したメッシュ気候図をそれぞれ示す。

3. 研究の方法

データすべてを使用して推定式の作成を行った場合の RMSE の値を図 4 に示す。ニューラルネットでは、学習回数が多い程 RMSE の値が減り、学習回数 2000 回以上では、重回帰分析よりも低い値を示している。

一方、図 5 の Leave-one-out cross validation の結果では、図 4 に示す内容とは対照的に、ニューラルネットでの学習回数が 100 回の場合が、1000 回の場合よりも RMSE が低い誤差を示した。また、ニューラルネットと重回帰分析との RMSE の比較では、すべてのニューラルネットの場合より重回帰分析が低い値を示した。このことは、データすべてを使用して推定式の作成を行った場合、必ずしも学習回数の増加が推定精度の向上につながるわ

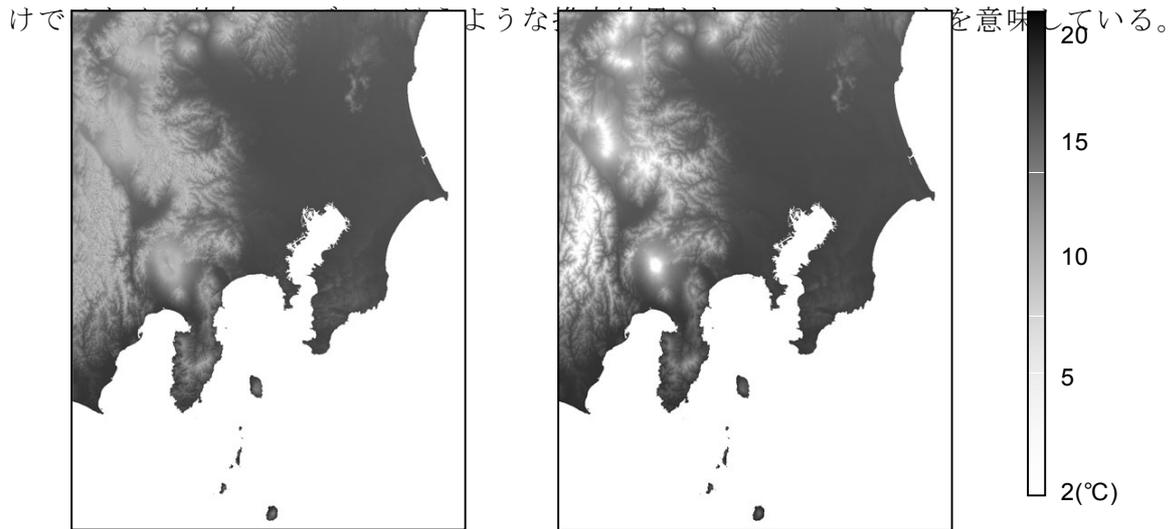


図 3 ニューラルネットを使用して作成したメッシュ気候図, 左図がニューラルネット, 右図が重回帰分析による結果を表す

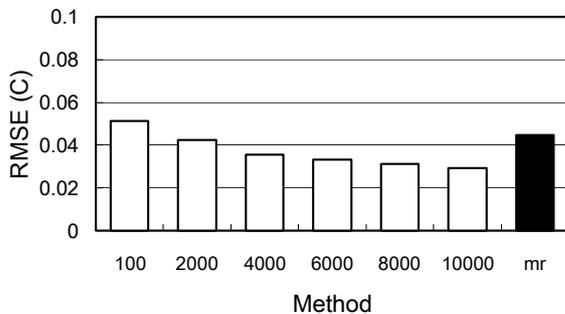


図 4 すべてのデータを使用して推定式の作成を行った場合の RMSE, 横軸の数字はニューラルネットの学習回数, **mr** は重回帰分析を表す

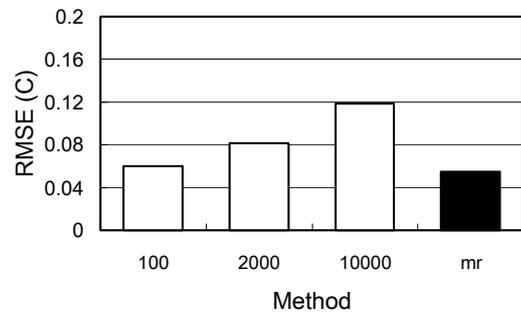


図 5 Leave-one-out cross validation を行った場合の RMSE, 横軸の数字はニューラルネットの学習回数, **mr** は重回帰分析を表す

したがって、推定手法として、ニューラルネットに、重回帰分析以上の精度が期待できるとはいえない。データすべてを使用して推定式の作成を行った場合の、各観測点での二乗誤差の分布を図 6 に示す。1000 回学習後のニューラルネットと重回帰分析の比較では、両者とも誤差の大きな点は比較的一致している。Leave-one-out cross validation の場合の各観測点での二乗誤差の分布を図 7 に示す。最も RMSE の小さかった 100 回学習後のニューラルネットと重回帰分析の比較では、ニューラルネットに誤差の大きな点があつたことがあり、これが RMSE の差の原因となっていることがわかる。また、図 6 と図 7 とを比べると、特に誤差の大きな地点が一致していることがわかる。

ニューラルネットでは、データすべてを使用した場合、学習回数が多い程 RMSE の値が小さくなり、Leave-one-out cross validation の結果では、学習回数が多い程 RMSE の値は大きくなった。このことから、学習回数が多いほど、学習に使用したデータに適した推定式が作成され、その推定式に学習データから離れた独立変数データを入力し推定を行っ

た

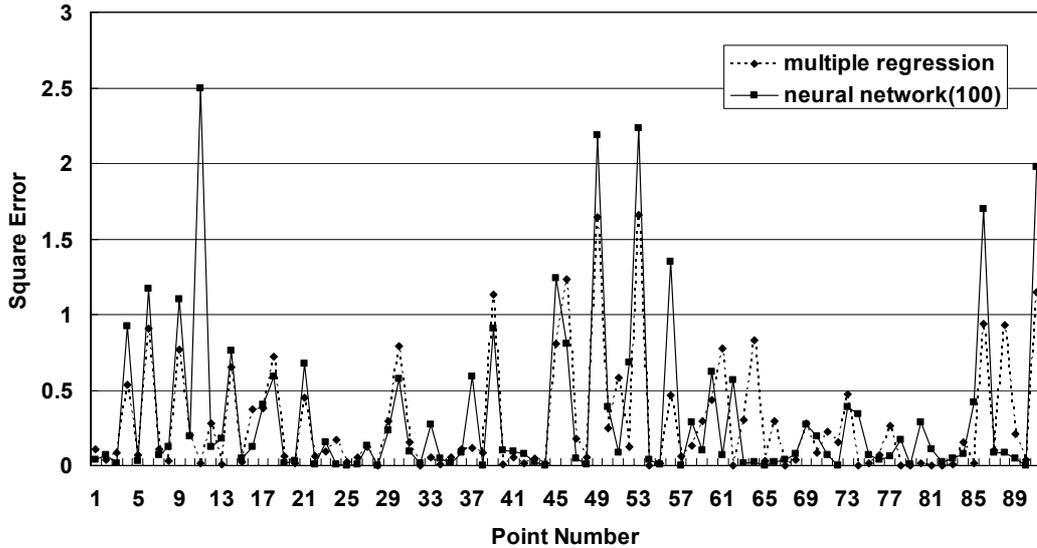


図 6 データすべてを使用して推定式の作成を行った場合の各観測点での二乗誤差の分布

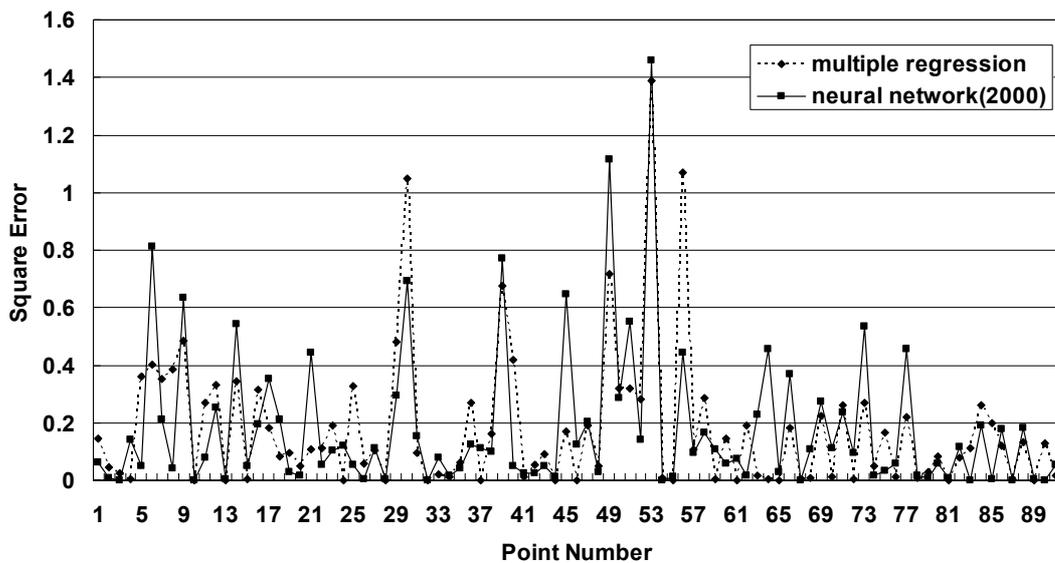


図 7 Leave-one-out cross validation の場合の各観測点での二乗誤差の分布

用いる場合、留意すべき点である。

どちらの手法を推定手法として用いるかは、従属変数と独立変数が、どの程度非線形の関係にあるかや、データ空間内でのデータの分布の状態などを考慮して決めるべきであろう。

4. 結論

メッシュ気候値作成時のニューラルネットワークと重回帰分析の精度比較のために、Leave-one-out cross validation 法を用いて分析を行った。その結果、ニューラルネットより重回帰分析が RMSE に低い値を示しており、推定手法として、必ずしもニューラルネットが、重回帰分析よりも精度が良いとはいえないことがわかった。ニューラルネットワー

クを用いた推定は数多く行われているが、その評価には注意を要するといえる。