

Prediction and interpolation of meteorological data using Recurrent Neural network

Suzuki Seiji Fujiki Nahomi

National Institute of Technology, Sendai College

1. はじめに

近年、ICTの農業等への活用が注目されており、様々なセンサを用いた遠隔でのデータの収集や解析が行われている。それらセンサには太陽電池を電源とすることが多く、気象状況や積雪等で発電できずデータを取得できなくなる問題がある。本研究では周辺で取得されたデータを基に対象のセンサデータの値を時系列データの解析によく用いられている Recurrent Neural network (RNN) にて予測し、欠落部の補間する手法の検討を行った。

本報告ではいくつかの気象データを対象に周辺観測所の違いによる誤差の変化と、RNNによる予測の有効性の確認、実際に欠落ある気象データの補間結果について報告をする。

2. 方法

RNNに予測対象の過去の気象データを教師データとし、周辺観測所の気象データを入力データとして使用し、学習することで、欠損率が大きく学習や解析などに使えないデータを周辺観測所から得た情報を基に予測する。

3. 実験

本報告では入力として参照する観測所の位置関係や数による予測結果の検証(実験1)と、RNNによる予測の有効性の確認(実験2)、実験対象農園(大倉ふるさと農園)の欠落のある気象データの補間実験(実験3)を行った。実験1に使用したデータは仙台高等専門学校(仙台高専)、仙台管区气象台(仙台)、山形地方气象台(山形)、福島地方气象台(福島)で測定された気象データであり、それらは全て気象情報提供サービスの「超高密度気象観測システム POTEKA NET(明星電気株式会社)」^[1]より取得した。取得した気象データは期間2018年5月1日0時10分~2019年4月30日23時50分の10分毎の気温、湿度、気圧、風速である。実験2では実験1のデータに風向を加えた。また、実験3では実験2で使用したデータに加えて、仙台高専付近に位置する大倉ふるさと農園に設置をしてあるセンサノードから取得した気象データを教師データとして使用した。

RNNについては、Gated recurrent unit (GRU)^[2]

を機械学習ライブラリの Keras を用いて使用した。構成は入力層の Neuron 数 4, 8, 12 個、中間層の GRU ブロック数 1000 個、出力層の Neuron 数 4 個、遡るステップ数 144 とした。また、学習の条件は学習定数を 0.0001、学習回数を 500 回、評価関数を平均二乗誤差、活性化関数を tanh と sigmoid 関数とした。

実験1では仙台高専の気象データが1ヵ月間ごとに欠落すると仮定して、欠落のおこる前の月を学習期間に設定し、その学習後に欠落を補間するための予測を行うことで予測性能を検証した。つまり5月のデータで学習をして、翌月である6月を予測し、次は6月のデータで学習をして、7月を予測するように1ヵ月ずつずらしていく形になる。予測の開始月は2018年6月として終了月は2019年4月とした。入力に用いた気象データの観測所による変化を調べるために、観測所を仙台、山形、福島、仙台+山形+福島と変化させて実験を行った。また、教師信号は仙台高専の気象データを使用した。

実験2では実験1で最も誤差が小さかった仙台+山形+福島を入力データとしたときの予測結果を使用した。この比較対象として、仙台高専に最も関係係数が高い仙台の気象データに定数倍をして補正したもの(以降、仙台補正データと表記)を使用した。定数はそれぞれ予測する前の月の平均気温、平均湿度、平均気圧、平均風速の仙台高専と仙台の比とした。しかし、風向はワンホットであるため、仙台の値をそのまま使用した。

実験3では大倉ふるさと農園のセンサノードで実際に欠落のあった2018年8月30日0時0分~2018年9月11日15時30分までの期間を予測して補間した。この時の学習データの期間は2018年8月1日0時10分~2018年8月29日23時50分までとした。

4. 結果と考察

実験1の各月での予測結果と仙台高専との誤差の平均と仙台高専に対する各観測地点の平均の相関係数のグラフを図1に示す。

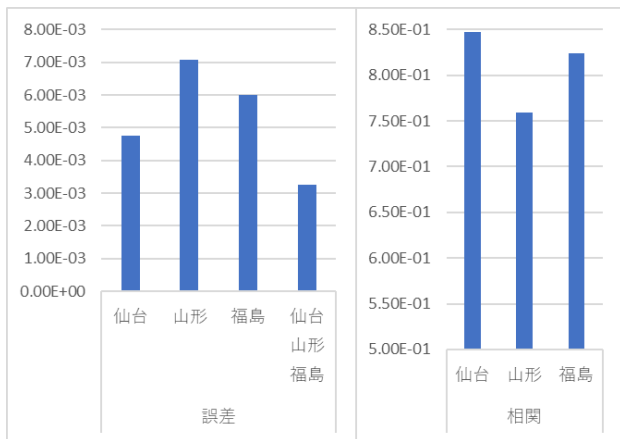


図1. 平均誤差と平均相関係数

誤差は仙台+山形+福島, 仙台, 福島, 山形の順で小さくなった. このことから観測地点が多いと誤差が小さくなるのがわかる. 加えて, 相関係数が高い観測地点では誤差が小さくなるのがわかった.

実験2の予測データと仙台補正データの気象データの平均二乗誤差のグラフを図2に示す. 風向は正解率[%]を示している.

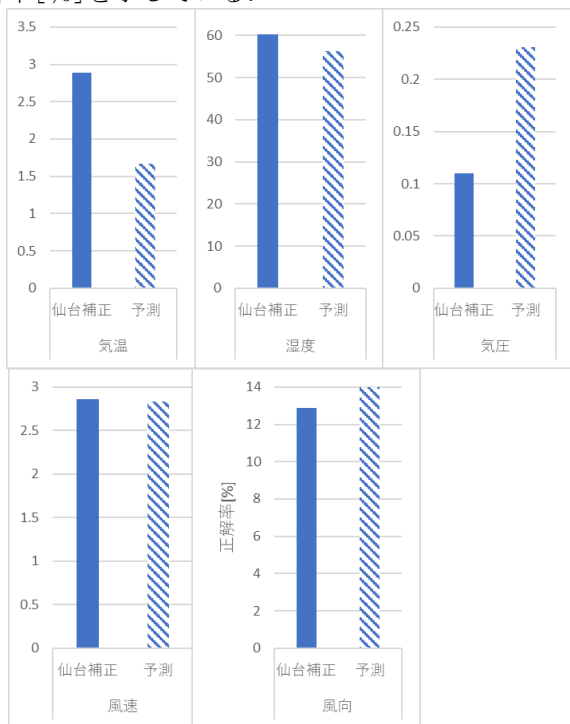


図2. 予測データと仙台補正データの誤差

気温と湿度での誤差の平均が予測データの方が小さく, 気圧では仙台補正データの方が小さく, 風速と風向ではそれぞれの平均値が非常に近い結果となった. このことから, 本研究の手法は気温と湿度の予測に特に有効であることがわかった. 気圧の誤差が仙台補正の方が小さい原因は仙台と仙台高専の気圧の相関係数が0.998と非常に高く, 補正によりさらに値が仙台高専

に近くなったことであると考えられる.

実験3の大倉ふるさと農園の気温を補間したグラフを図3に示す.

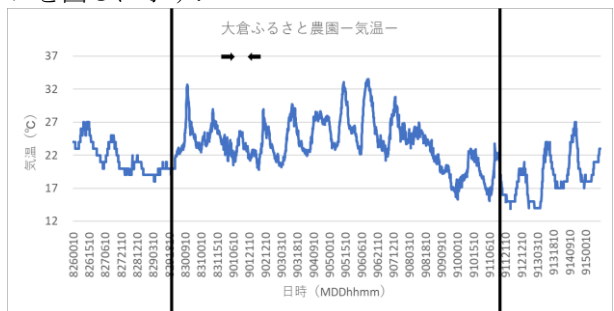


図3. 補間をした大倉ふるさと農園の気温グラフ

黒い線で囲まれている部分が補間期間となっている. このグラフを見てみると違和感なく補間ができていることがわかる. また, 補間期間内で断片的に残っていた2018年9月1日9時10分~18時50分(図3矢印間)の大倉ふるさと農園の実測値と予測値の誤差の平均は2.3[°C]であった. 気象庁の翌日の最高気温予報の誤差は2018年9月の東北地方で約1.7[°C]であり[3], それほど大きな誤差ではないことが分かる. そのため, RNNを用いて気象データの補間する方法は有効であると考えられる.

5. まとめ

今回の実験では周辺観測所の違いによる誤差の変化と, RNNによる予測の有効性の確認, 実際に欠落ある気象データの補間を行った. 入力データに用いる気象データは補間対象の気象データと相関係数が高く, 複数の観測所のものを用いると予測精度が向上することが分かった. RNNによる予測は特に気温と湿度の予測に有効であることがわかった. また, 気圧に関しては相関係数の高い周辺の気象データを対象の気象データとの比で定数倍した方が誤差が小さくなることがわかった. 実際に欠落のある大倉ふるさの農園の気象データの補間では, 小さい誤差で違和感のない補間をすることができた.

参考文献

[1] 明星電気株式会社, “超高密度気象観測システム POTEKA NET”
<http://www.potekanet.com/index.php>,
 (2020年1月9日時点)

[2] Junyoung Chung, Caglar Gulceche, KyungHyun Cho, Yoshua Bengio, “Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling”, arXiv:1412.3555, Dec. 2014

[3] 気象庁, “天気予報の精度検証結果—地方区別グラフ”
<https://www.data.jma.go.jp/fcd/yoho/kensho/tenkiyoho.html>,
 (2020年1月9日時点)