

統計的日射予測における地点間隣接関係の活用

比護貴之^{†1}

概要: 本稿は、数時間先の日射を統計的に予測することを、数百の学習データを用いて、数万の説明変数から、数万の目的変数を予測する回帰問題とし、このためのモデル選択手法として、地点間の隣接関係を活用する手法について議論する。既存手法であるマルコフ確率場を用いた方法では問題が生じ、その解決方法や、技術課題を明らかにする。

Utilizing Spatial Adjacency for Statistical Irradiation Forecast

Takayuki Higo^{†1}

1. はじめに

電力システムの運用では、急速に導入が進んでいる太陽光発電 (PV) の出力変動を事前に予測することが課題となっている。PV 出力変動を予測するには、その要因となる日射変動を予測する必要がある。しかし、現状の気象予測技術では系統運用に必要な日射予測精度が得られない。

この日射予測に対しては、統計的 (機械学習的) なアプローチの研究例が少ないため、改善の余地があると考えられる。統計的に日射を予測する問題は、回帰問題である。ただし、この回帰の問題は、たとえば、6 万地点の 1 年分のデータを使う場合は、365 の学習データに対して、説明変数 (各地点の予測実施時点の日射) と目的変数 (各地点の予測実施時点の日射) の数が 6 万もあるため、モデル選択 (過学習の抑制) が課題となる。

実用的なモデル選択手法はあるが、それらは限られたデータだけからのモデル選択手法である。更なる高精度化に向けては、データ以外の情報として、地点間の隣接関係を用いる方法が考えられる。たとえば、前記の 6 万変数は、6 万地点に対応しており、各地点に対しては隣接地点を想定できる。ある地点の気象は、隣接地点の気象と類似しており、また、ある地点の気象の変化は、隣接地点の気象状況によって引き起こされる (たとえば、隣接地点の雲が移動してくる等)、という仮定をモデル選択に導入できれば、精度向上につながる可能性がある。

本稿の目的は、日射予測の回帰問題において、隣接地点関係を適切に活用する手法の有用性と課題を明らかにすることにある。隣接関係を活用する手法としては、ガウシアンマルコフ確率場 (GMRF) [1]がある (GMRF はガウシアングラフィカルモデルと呼ぶこともある)。しかし、事前検討により GMRF には問題があることがわかった。問題の原因は十分に明らかではないが、目的変数が有する回帰誤差を考慮していないことが原因の可能性があり、これを克服

する手法を本稿では新たに提案する。実データを用いた数値実験を行い、提案手法の有用性および日射予測における隣接関係活用に向けた課題を考察する。

2. 統計的日射予測の問題設定

本稿で扱う統計的日射予測について説明する。現在、衛星画像を用いて、1km 空間解像度の日射強度データを得ることができる。そこで、たとえば、朝の 9 時に、3 時間先である 12 時の日射を予測する場合、説明変数は 9 時の各地点の日射、目的変数は 12 時の各地点の日射とする。扱う地点数は検討が必要だが、関東エリアを想定し、330km × 230km 格子の各格子地点を扱うことを本稿の目標とする

統計的日射予測は回帰問題であり、回帰式 (予測式) は、回帰係数の行列 B を用いて次のように書ける。

$$xB = y \quad (1)$$

ここで x は予測実施時点、 y は予測対象時刻の各地点の日射を表す行ベクトルである。

3. 提案手法

提案手法では、説明変数と目的変数それぞれに対して、隣接する変数を説明変数とした回帰式を作成する。回帰係数は次の等式を近似的に満たすように決める。

$$(x \hat{y})B = (x \ y) \quad (2)$$

ここで、 x, y, B は式(1)と同じであり、 $(x \ y)$ は x と y を行方向に連結した行ベクトルである。 \hat{y} は、 y に回帰式の誤差を反映したものであり、回帰係数 B を、学習データに適用した値である。 \hat{y} によって B が決まり、 B によって \hat{y} が決まることから、交互に \hat{y} と B を計算することを繰り返すことで、 B を決定する。

回帰係数行列 B の推定では、隣接関係のある変数に該当する要素以外をゼロとするために、列毎に最小二乗法で決定する。ただし、最小二乗法は、すなわち、重回帰の計算

^{†1} (一財)電力中央研究所
Central Research Institute of Electric Power Industry

だが、提案手法では厳密には Ridge 回帰の形に修正して計算する。この理由は、前記の繰り返し計算を行うと、 \hat{y} に回帰誤差が反映され、この結果、一部の \hat{y} が情報を持たなくなるためである（たとえば、共分散行列がランク落ちする）。Ridge 回帰の正則化パラメータとしては、事前に決めた 10 個の値から、交差検定 (LOOCV) の誤差を最小化するものを用いる。

回帰係数 B が決まれば、 x を既知変数、 y を未知変数とし、次の線形方程式から y を求めることができる。

$$(x \ y)(B - I) = 0 \quad (3)$$

ここで I は単位行列である。

4. 数値実験

本稿の数値実験で用いる日射データは、衛星画像に基づく関東地域の日射推定値[2]であり、使用するデータの期間は 2016 年 3 月 1 日からの 1 年分である。また、日射としては後 30 分間平均値を使用する。

隣接関係については、各地点は東西南北に隣接する 4 地点と隣接しているとする。また、時間方向については、1 時刻前の同地点と東西南北の 4 地点と隣接しているとする。

4.1 実験 1

330km×230km の地域 (330×230 地点) の 9 時の日射を対象とし、一部の地点 (347 地点) の日射を既知とし、残りの日射を予測する内挿の問題を考える。つまり、この問題は時間発展を考慮していない。手法としては、GMRF と提案手法の他に、ユークリッド距離に基づく k-NN を用いる。k-NN のパラメータ K は、事前検討から高精度であった 3 とする。

各手法の予測誤差を図 1 に示す。誤差は、絶対値誤差を地点間で平均した後、年間で平均した値である。GMRF の誤差が大きいことが確認できる。一方、誤差が一番小さいのは k-NN であり、これと比べて提案手法の誤差は、11% 大きい。この結果から、GMRF の適用には前述したとおり問題があり、この問題は提案手法により一部改善し、k-NN に近い精度となる。なお、k-NN は時間方向の予測に適用させることは単純でないが、提案手法は隣接関係を使って時間方向の予測にも拡張できる。図 2 にある日の予測結果を示す。白い部分は日射が強いことを示す。

4.2 実験 2

70km×70km のエリアを対象に、9 時の時点の日射を使用して、10 時の時点の日射を予測する。提案手法では、9 時と 9 時半、9 時半と 10 時の日射が隣接関係にあるとして、予測する。また、比較する手法として、重回帰を用いる。重回帰の説明変数は、予測対象地点とその地点の東西南北に隣接する 4 地点である。

結果を図 3 に示す。提案手法の地点間で平均した絶対値誤差の年間平均値は、重回帰より 2% 大きい。一方で、提案

手法の誤差の年間の中央値は、重回帰より 10% 小さい。これは、提案手法は通常時は高精度な予測をできるが、まれに大きな誤差が発生する場合があると解釈できる。この大きな誤差の要因としては、当日の気象状況や雲の種類の可能性がある。

5. 考察

数値実験から、GMRF に対する提案手法の有用性を確認した。一方で、1 時間先予測の予測精度について、重回帰と比較した提案手法の有用性が一部認められるものの、実用上はより明確な改善が必要である。提案手法については、Ridge 回帰の使用の意味の解明や、収束の高速化、適切な隣接関係について課題がある。また、GMRF の問題についても、その要因を解明する必要がある。

謝辞 本稿の実験では、電力中央研究所の橋本篤氏が作成した日射推定値データを使用させていただきました。ここに感謝の意を表します。

参考文献

- [1] Hastie, T. et al.. The elements of Statistical Learning. Springer Series in Statistics Springer New York Inc., 2001.
- [2] 橋本篤, 宇佐美章, ひまわり 8 号を用いた日射量推定・予測システムの開発, 電力中央研究所報告, 2017.



図 1 予測誤差の比較 (実験 1)

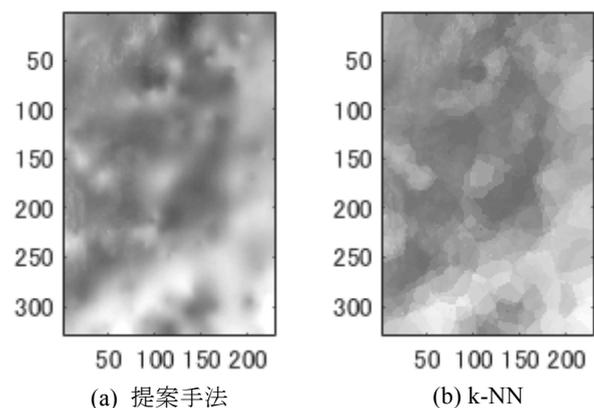


図 2 予測結果の例

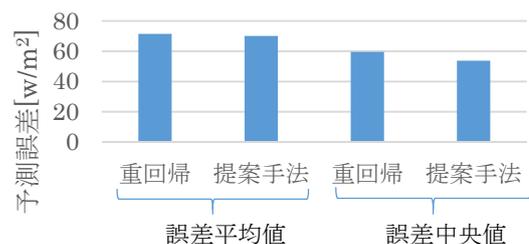


図 3 予測誤差の比較 (実験 2)