

GPV と発電データを利用した 機械学習のモデルアンサンブルによる予測技術

PV output forecasting by model amusable of machine learning
methods using GPV

PV-4cast

大関 崇^{*1}・大竹秀明^{*1}・宇野史睦^{*1}・海崎光宏^{*1}・João Gari da Silva Fonseca Júnior^{*2}

1. はじめに

2018年7月に第5次エネルギー基本計画が閣議決定され、太陽光発電（以下、PV）のさらなる導入拡大が期待されている。政府が現時点で示している2030年のエネルギーミックスにおけるPVの目標値は64GWであるが、国内におけるPVの累積導入量はすでに2017年末時点で50GW（DCベース）の達成間近であり、2030年を待たずこの目標値を達成すると考えられる。しかしながら、導入するだけでなく、長期的にエネルギーインフラとして定着することがPVの主力電源化の実現には必要不可欠である。

それを実現するためには、需給運用において利用されているPVの発電予測技術のさらなる予測精度向上が必要とされている⁽¹⁾。

このような背景の中、東京電力ホールディングス株式会社と北海道電力株式会社により、革新的な手法や新しいアプローチの発掘を目的に、北海道地域での太陽光発電量予測をテーマとした太陽光発電量予測コンテスト『PV in HOKKAIDO』を共同で開催された⁽²⁾。本稿では、筆者らがコンテストに参加した際の発電予測モデルなどについて紹介する。

2. PV 発電予測モデル

可能な限りシンプルな予測モデルにすることを目的として、発電実績データ（発電電力量）と気象庁配信のMSM-GPVのみを利用するモデルとした。特徴としては、追加の日射データ等が不要、方位、傾斜等の情報が不要、他のエリアにおいても汎用的に利用可能（他エリアの発電実績データは必要）である。また、本コンテストの対象が北海道エリアで

あることから、積雪対応について考慮するモデルとした。

2.1 PV 発電予測モデルの概要（ベースモデル）

前日予測であるため、数値予報データを入力とし、発電電力量を出力とする学習モデルとした。

予測モデルは、前日15時配信される数値予報データ（MSM-GPV）と大気外日射量を入力すると、翌日の1日分の発電電力量を出力する学習モデルである。学習と予測は1ヶ月単位で行い、学習データは過去2年分の同月データとした。発電電力量は30分値であるため、数値予報データ（1時間値）を線形補間し30分値とした。

学習モデルは、Support Vector Machine（以下、SVM）、Random Forest（以下、RF）、XGBoost（以下、XGB）の3つの機械学習モデルを利用し、それぞれの予測結果の算術平均を最終的な予測値とした。入力パラメータの選定や学習方法については、FonsecaらがGPVとSVMを利用したモデルを参考とした⁽³⁾。

予測計算はR言語により行ない、SVM、RF、XGBの計算はそれぞれ、e1071、randomForest、caret（xgbTree）パッケージを用いた。各パラメータについては、SVMはe1071パッケージのtune.svm（）でパラメタの範囲を指定せずチューニングし、RFはrandomForestパッケージのtuneRF（）でチューニングした。XGBはcaretパッケージを使い、検証するパラメタの数はパラメタ毎に4（tuneLength）とし、クロスバリデーションによりチューニングを行った。

また積雪対応として、MSM-GPVの気温と相対湿

^{*1} 産業技術総合研究所

^{*2} 東京大学生産技術研究所

度から降雪量を推定，降雪量の積算により積雪量を推定し⁽⁴⁾，入力データとして利用した．降雪量は30分毎に計算し，積雪量は降雪量を積算し，前日の積雪量を説明変数とした．

$$T > T_c : \text{Rain} \quad \text{式 (1)}$$

$$T \leq T_c : \text{Snow} \quad \text{式 (2)}$$

ここで， $T_c \sim 7.7 - 6.6 \text{ RH}$ ， $0.4 < \text{RH} \leq 1.0$ (T:地上気温 (°C)， T_c :雨雪閾値，RH:相対湿度)

※雨量を降雪量とする

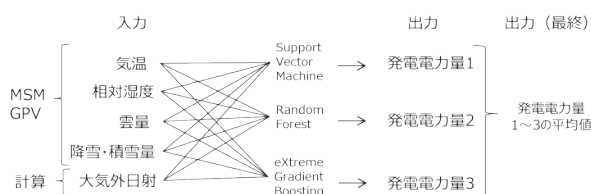


図1 予測モデルの概要

3. 利用したデータ

発電実績データは，北海道エリア内で系統に接続されている太陽光発電所のうち，コンテスト主催者が予め指定する15箇所，合計407.3 MWであり，エリアが大きく2箇所に分割され，それぞれのエリア（S1，S2）の合計発電電力量を利用した．

学習時の発電実績データは，S1，S2をそれぞれ利用し，気象庁配信のMSM-GPVは，各サイト最寄り格子点データをすべて利用した．なおGPVデータは，京都大学においてアーカイブされているデータを利用した⁽⁵⁾．

初期値は，日本時間の15時配信（JST15），使用する変量は，気温，相対湿度，時間降水量，雲量（上層・中層・下層）とし，降水量は，降雪量の計算のみに使用した．

予測期間は，2018年1月1日～2019年1月31日とし，雪考慮期間は1-3，12月とした．

計算環境は，汎用的なPCを利用し，Windows 10 Pro，CPU Intel (R) Xeon (R) CPU E5-2667 2.90 GHz 2.90 GHz（2プロセッサ）※プログラムは並列化せず計算，メモリ128 GB ※実際の使用メモリは1GB未満を利用し，プラットフォームはR 3.5.2 (Platform: x86_64-w64-mingw32/x64 (64-bit)) を利用した．

4. 入力データ変更の検討

2.1節のベースモデルに加えて，入力にMSM-GPVの予測日射量および前日発電電力量を追加し

た場合の影響を調査するために，以下のそれぞれについて検討を行った．

① MSM-GPVの予測日射量を追加した場合

② 前日発電電力量を追加した場合

③ MSM-GPVの予測日射量+前日発電を追加した場合

ここでベースモデルは，過去2年分の同月により学習しているが，ここでは日射量データが2年分ないため，過去60日学習を利用した．（2月5日に対する予測から利用）また，ベースモデルはS1，S2に分けて予測したが，ここでは全発電電力量を目的変数とした．予測期間は，2018年1月1日～2019年1月31日（雪考慮期間：1-3，12月）とし，過去60日学習では12月1，2日を除いた．前日の発電電力量は，前日17:00までのデータとした．

各ケースの説明変数の条件などについて図2に示す．

ID	略称	目的変数	説明変数								学習期間	
			気温	湿度	雲量	降雪	積雪	大気外日射	日射	前日発電		
A1	ベースモデル	S1, S2の発電電力量	○	○	○	○	○	○	○	-	-	同月 (2ヵ月)
B1	過去60日	全発電電力量 (S1 + S2)	○	○	○	○	○	○	○	○	-	過去60日
B2	過去60日 (雪量なし)		○	○	-	○	○	○	○	○	-	
B3	過去60日 (日射なし)		○	○	○	○	○	○	○	-	-	
C1	前日発電	全発電電力量 (S1 + S2)	○	○	○	○	○	○	○	-	○	同月 (2ヵ月)
C2	ベースモデル (全発電)		○	○	○	○	○	○	○	-	-	
C3	前日発電 (雪量なし)		○	○	○	-	-	○	-	-	○	
D1	過去60日 (前日発電)		○	○	○	○	○	○	○	○	○	過去60日

図2 各ケースの概要

5. 予測誤差の結果

5.1 ベースモデルの結果

ベースモデルの月ごとのRMSEを図3示す（分

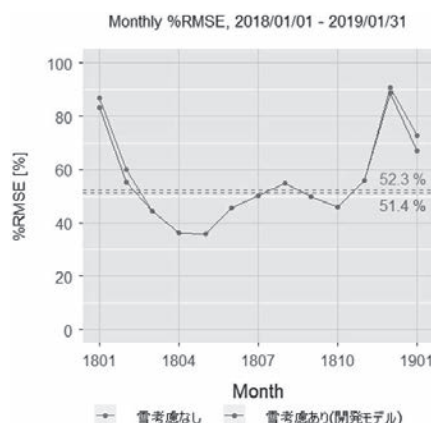


図3 ベースモデルの予測誤差結果

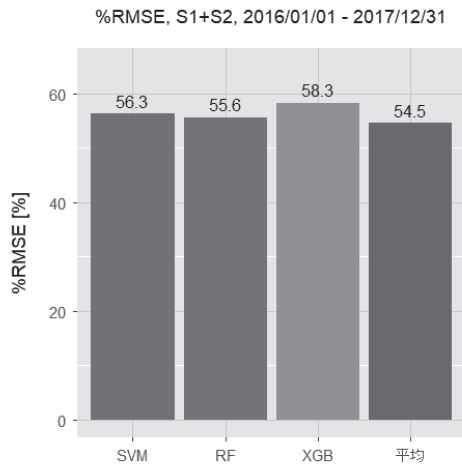


図4 3つの機械学習の比較結果

母は月ごとの積算発電電力量)。冬季における予測誤差が悪い傾向であった。降雪量と積雪量データを入力に追加し、予測精度は1%程度改善した。また、3つの機械学習の結果を図4に示す。それぞれのモデルを平均することにより誤差低減している。

5.2 入力データ変更の検討の結果

入力データを変更した検討結果として、に MSM-GPV の日射量を追加した結果を図5に示す。予測誤差は日射+雲量が最も小さく、次に日射量データ

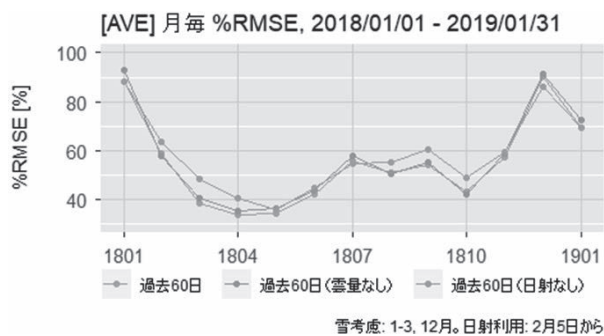
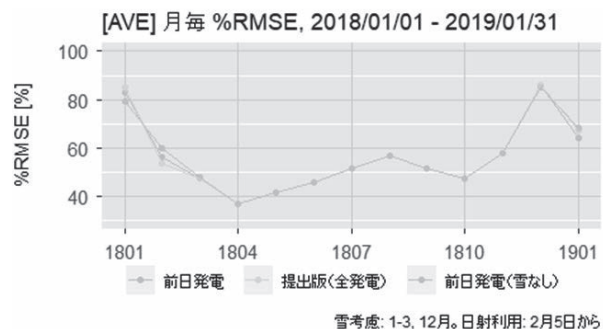
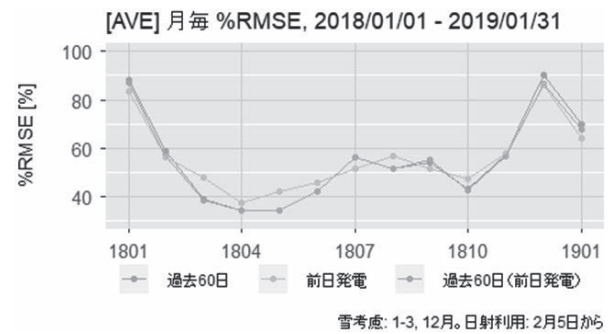
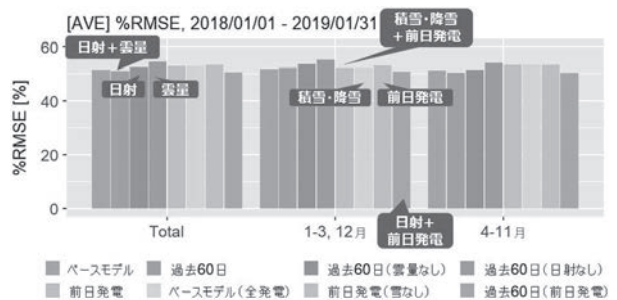
図5 日射量を追加した場合
両方あり・日射のみ・雲量のみ図6 前日発電を追加した場合
両方あり・雪のみ・発電のみ図7 日射量+前日発電を追加した場合
過去60日(赤)と過去60日+前日発電(灰色)

図8 パラメータ変更による予測誤差結果

のみが小さく、雲量データのみが最も大きい結果となった。

前日発電電力量を追加した結果を図6に示す。予測誤差低減効果は小さい結果となり、SVMでは微増となった。積雪データの追加と前日発電電力量の追加を比較した結果、積雪データを追加する方が、前日発電電力量を追加するよりも誤差が小さい結果となった。また、XGBはRFより高精度であった。

予測日射量+前日発電電力量を追加した結果を図7に示す。前日発電電力量を加えたことで、冬場に予測誤差を低減できた。特にXGBの改善が大きかった。

これらの総合の結果比較を図8に示す。ここで図3,4で示すベースモデルの予測誤差が異なるのは、入力データの違いによる予測誤差比較が目的であるため、評価期間を利用可能なモデルにあわせているためである。

6. 考察とまとめ

可能な限りシンプルな予測モデルを目的として、発電実績データ(発電電力量)と気象庁配信のMSM-GPVのみを利用するモデルとした。コンテストにおいては予測誤差の上位には入れなかったが、MSM-GPVのみを利用する点で評価をいただいた。デメリットとしては、機械学習を利用しているため、ブラックボックスモデルのため、予測誤差の要因の

説明が難しい点が挙げられる。

今後の予測誤差低減の可能性については、以下が考えられる（特に冬季の積雪について）。

- ・地上気象官署や衛星観測の日射や積雪などの気象データの活用。
- ・個別発電所データの利用。
- ・融雪モデルの導入。
- ・複数予測と識別器・ブレンディングの利用。

アレイ面への積雪状況は、システム依存性が強いと考えられるため、今回は利用できるのデータがエリア合計のみであったため、個々の発電所データがある場合は、日射と発電の関係からアレイ面への積雪の程度が一定程度別可能であると考えられる。また、気象側では融雪モデルなどもあるため⁽⁵⁾、これらの物理モデルからのパラメータ導入も検討が可能である。ただし、上記モデルでは、追加の気象データ等が必要となるため、今回のモデルがシンプルという利点は失われる。

最後に、コンテストを通じて様々な手法の比較が行われ、それぞれ特徴と予測誤差の差異が見られた。これは、さらなる予測精度向上の可能性が残されていることを示唆しており、さらなる研究開発の必要

性を再認識した。

このような機会を与えていただいたことに、関係者の皆様に感謝する。

7. 参考文献

- 1) 第19回 総合資源エネルギー調査会 電力・ガス事業分科会 電力・ガス基本政策小委員会資料
- 2) 太陽光発電量予測技術コンテスト『PV in HOKKAIDO』: <https://cuusoo.com/projects/50369> (2020/01/09 アクセス)
- 3) Fonseca Jr Joao Gari da Silva, Oozeki Takashi, Takashima Takumi, Koshimizu Gentarou, Uchida Yoshihiza, Ogimoto Kasuhiko, "Use of support vector regression and numerically predicted cloudiness to forecast power output of a photovoltaic power plant in Kitakyushu, Japan", Progress in Photovoltaics: Research and Applications (2011)
- 4) 京都大学 GPV アーカイブ: <http://database.rish.kyoto-u.ac.jp/arch/jmadata/data/gpv/original/>
- 5) 近藤純正, "環境の気象学－地表面の水収支・熱収支", 朝倉書店 (1994)