Solar Power Forecasting Using a Stacked Ensemble of Strong Learners

Christopher Dallosta

12/12/2020

Introduction

In 2016, 196 members of the United Nations Framework Convention on Climate Change signed the Paris Agreement setting a goal to increase renewable energy's market share to 20% (Paris Agreement). This accord was the first global agreement to begin divesting in fossil fuels and looking forward towards a future of renewable energy. More recently, during the 2020 elections, Nevadans voted in favor of Question 6 which initiated a constitutional amendment to acquire 50 percent of electricity in Nevada from renewable resources by 2030 (Nevada Question 6). These two events indicates a major shift in civil and public opinion away from traditional fossil fuels and towards renewable energy such as photovoltaic power, henceforth to be referred to as PV.

Up until 2015 solar panel efficiency has hovered below 20 percent. Recently solar panel technology has experienced a renaissance as a group of scientists in 2015 developed a solar cell capable of 44.5% efficiency (PV efficiency). Continually, the cost of large scale PV panels has decreased by 73% from 2010 to 2017, making PV an efficient and cost effective renewable energy solution. However, despite these advancements PV still has a number of challenges.

PV power output is highly variable and contingent on weather variables (Carrera et al). As a result, a balancing act of PV power production and consumer consumption is executed by utility companies daily. In order, to ensure PV as a profitable renewable energy option accurate PV power output forecasting has become an important function. Both time series and machine learning techniques have been used to predict PV power output to varying degrees of accuracy. Advances in this field continue to be developed and with existing models continually improved upon. At each model's core, PV forecasting models rely on the relationship between weather variable inputs and PV power output. High levels of model accuracy can be achieved by creating a model that mirrors this relationship

Artificial neural networks recently have been proven to be the frontrunner of models that can achieve a high level of forecasting accuracy. This goal of this paper is to compare the results of an Artificial Neural Network, a Support Vector Machine, and a Gradient Boost Regression Tree to a stacking ensemble containing all three of the aforementioned machine learning models. I will be using data from the European Centre for Medium-range Weather Forecast in order to create a PV forecasting model.

Literature Review

Data

An analysis of a variety of studies identified that most PV forecasting models use three formats of data. Studies either used weather forecasts, weather observations or a combination of the previous two variables (Carrera et al). Although weather forecasts and weather observations are prone to collection errors, it is to be noted that weather forecast data is subject to errors in the forecasting model as well. Carrera et al performed a study to compare the accuracy of machine learning models using all three types of data. They found that although weather forecasts and a combination of weather forecasts and observed weather performed similarly, a combination of the two data types consistently produced higher levels of model accuracy regardless of the machine learning model used.

Pre Processing

Preprocessing of data typically is one of the most important steps to ensuring consistent model performance. A recent study performed by Coimbra and Pedro on forecasting techniques for solar power production found that model accuracy tends to decline when using the same model for different climatic regions. However, when a model is fit to a specific region then model accuracy improves. (Coimbra and Pedro). Furthermore, Kubby and O'Leary, in their study 'Feature Selection and ANN Solar Power Prediction', found that controlling for the time of day helped to improve model accuracy. They found that there are four distinct time frames that power output can be categorized into. Night and Day (sunny days) are typically rather constant at a high power output during the day or null power output at night. Sunrise and Sunset on the other hand typically contain the highest error rate due to their volatility.

In order to reduce noise and/or dimensionality Congedo et al studied the effects of using Principal Component Analysis, Wavelet Decomposition, and a combination of the two methods on model performance. Principal Component Analysis (PCA) is a statistical technique that takes into account uncorrelated and/or redundant data to reduce the number of dimensions in a dataset, whereas wavelet decomposition (WD) uses time-scale signal analysis to deal with solar irradiance fluctuations, thereby reducing the noise. Although each preprocessing technique tested performed similarly well, Congedo et al found that performing PCA on a dataset and then performing WD on the dimensionally reduced data provided the largest increase in model performance.

Model Selection

Congedo et al leveraged a variation of the Support Vector Machine (SVM) function in cohesion with their data preprocessing techniques to achieve their high model accuracy levels. A Support Vector Machine is a method that "tries to maximize the distance between the separating hyperplane and the support vectors within a threshold value" (Carrera et al).

Gradient Boosting Regression Trees (GBRT) leverages the power of a large number of weak performing models, such as a decision tree, to create a singular well performing model. Weak learning models are sequentially fitted to a dataset. On each iteration the previous model is used as a starting point and improved upon in order to decrease the overall model's bias. Conde and Isaksson in their study 'Solar Power Forecasting with Machine Learning Techniques' tested a multitude of machine learning models to identify techniques that would best forecast solar power output. Consistently, both their ANN and their Gradient Boost Regression Tree model performed at a higher level than the other models tested (Conde and Isaksson).

Artificial Neural Networks (ANN) mimic the physical structures of the brain. An ANN consists of a total of three layers, the first being the input layer. Each node of the input layer takes in a singular data source and passes that data to the next layer, the hidden layer. The hidden layer can contain any number of nodes, or neurons, of which takes in the input, applies an activation function to the input, and passes the data to the next layer. There can be any number of additional hidden layers. The last layer is the output layer which compiles the data from the previous layers and outputs a predicted value. This process is repeated and the weights between each node is adjusted to reduce the error of the prediction. Chen et al. used an Artificial Neural Network to perform long term, up to 100 hours, power output forecasting using weather forecast variables. Although, overall, their model performed well, they found that rainy days caused a high amount of model error in contrast to the high levels of accuracy achieved on cloudy and sunny days (Chen et al).

Gonzalo Martinez-Munoz, in his study "Sequential Training of Neural Networks with Gradient Boosting", leveraged both the benefits of Gradient Boosting with the high performance of Artificial Neural Networks. Martinez-Munoz iteratively trained shallow neural networks from the residuals of the previous model. He then combined each trained model to create a single neural network with one hidden layer. He found that this training method provided a method with a lower prevalence of overfitting and performed more favorably than a standard neural network (Martinez-Munoz).

Methodology

Data

The data used for this project was from the European Centre for Medium-range Weather Forecasts (ECMWF). The data included 12 weather variables that were collected every hour on the hour. The hourly measurements were collected between April 1st, 2012 and July 1st, 2014. The data was collected from the location of three solar power plants of which were located in different geographical and climatic regions in Australia. A description of the weather variables can be found in Figure A. The dataset was then split into training data, of which spanned from 04/01/2012 to 07/01/2013, and a test dataset, which spanned from 07/01/2013 to 07/01/2014. The data was then split into three separate zones for each solar power plant location. Lastly, the power output were then shifted backwards 24 hours so that each independent variable would be trained on the power output that occurred 24 hours in the future.

Variable id.	Variable name	Units	Comments
078.128	Total column liquid water (tclw)	$kg\ m^{-2}$	Vertical integral of cloud liquid water content
079.128	Total column ice water (tciw)	$kg\ m^{-2}$	Vertical integral of cloud ice water content
134.128	Surface pressure (SP)	Pa	
157.128	Relative humidity at 1000 mbar (r)	%	Relative humidity is defined with respect to saturation of the mixed phase, i.e., with respect to saturation over ice below −23 °C and with respect to saturation over water above 0 °C. In the regime in between, a quadratic interpolation is applied.
164.128	Total cloud cover (TCC)	0–1	Total cloud cover derived from model levels using the model's overlap assumption
165.128	10-metre *U* wind component (10*u*)	$m\ s^{-1}$	
166.128	10-metre *V* wind component (10*v*)	$m\ s^{-1}$	
167.128	2-metre temperature (2*T*)	K	
169.128	Surface solar rad down (SSRD)	$J\ m^{-2}$	Accumulated field
175.128	Surface thermal rad down (STRD)	$J\ m^{-2}$	Accumulated field
178.128	Top net solar rad (TSR)	$J\ m^{-2}$	Net solar radiation at the top of the atmosphere. Accumulated field
228.128	Total precipitation (TP)	m	Convective precipitation + stratiform precipitation (CP + LSP). Accumulated field.

**Figure A – ECMWF weather variable descriptions

Pre-Processing

As indicated by Kubby and O'Leary time is an important factor in relation to potential solar power output. As a result, a feature indicating the hour of day was added to the dataset to allow the machine learning algorithms to train on the time of day. As seasonality has a high potential of affecting weather variables indicator variables indicating the season of the year in Australia were also added.

In order to get an idea as to the spatio-temporal correlation between the independent variables and the target variable a visual inspection via graphical representation was performed. Continually, a correlation matrix was constructed to identify potential existing relationships between independent variables. Lastly, the gradient boost regression tree algorithm performs splits that minimizes its cost function. In this manner the GBRT algorithm relies on the most important features to create a regression tree. The training dataset was passed through the GBRT algorithm to take advantage of this process. The split criteria was then extracted from the model and presented in graphical format to identify features that were identified as important by the GBRT algorithm.

The aforementioned data analysis provided a number of insights.

1. A low rate of change in relative humidity was a potential indicator for low power output. In order to capture this relationship a variable containing the natural log of each observation of relative humidity was created.
2. The GBRT algorithm placed a high level of importance on the surface thermal radiation weather variable. Continually, a visual analysis showed a non-linear relationship with power output. In order to capture this relationship a variable containing the cube of each observation of surface thermal radiation was created.
3. Intuition provided that the difference between top net solar radiation and surface solar radiation would indicate potential obstructions. As a result a variable differencing top net solar radiation and surface solar radiation was created. A visual inspection of this variable and the target variable indicated a potential relationship between the rate of change and power output. As a result, in order to capture this relationship the variable was replaced by natural log of the difference of top net solar radiation and surface solar radiation.
4. I wanted to test the relationship between seasonality and the ground temperature on power output. As a result, four variables interacting the season and 2-metre temperature was created.

Null values were replaced with zeros and non-categorical variables were then normalized so that their means were centered on 0 with a min of -1 and a max of 1.

In order to further hone in on the optimal set of features the forward stepwise algorithm was used. Forward stepwise starts by modeling without any features. It then tests the addition of each feature by calculating the performance of the feature using root mean squared error and adds the feature that had the highest performance value. It repeats this process until all features have been added and then goes back and selects the highest performing subset of features. This process paired with practitioner intuition indicated that 7 of the 24 variables needed to be removed in order to increase model performance.

Model Selection

Extensive literature review revealed a trend of Artificial Neural Networks, Gradient Boost Regression Trees, and Support Vector Machines as generally producing good overall results. However, during testing I noticed that each model tended to have unique biases. Continually, the work of Gonzalo Martinez-Munoz in creating a Gradient Boosted Artificial Neural Network provided inspiration for using an ensemble method. As a result, I decided to test a hypothesis that combining the results of the SVM, ANN and GBRT models using an ensemble method would produce a higher performing model.

Low level testing of AdaBoost, voting regressors, and stacked generalization models indicated that using a stacked generalization model would provide the best results. The stacked generalization models used for this model stacks the predictions of each individual estimator together. The predicted values of the stacked models are then used to train a final estimator that attempts to reduce the error of the stacked models and produces a final set of predicted values. The final estimator used for this model was the default RidgeCV estimator. The benefit being that the final estimator learns individually how to best combine the predictions from the multiple stacked models.

Hyper Parameter Tuning and Model Fitting

Each model was trained in turn for each individual zone. In order to identify the optimal hyperparameters the gridsearch CV algorithm was used to exhaustively test different combinations of hyperparameters. Due to the high computational requirements of this method, particularly for the ANN, integer parameters were tested at larger step intervals than desired. Model performance during hyperparameter tuning was evaluated using the root mean squared error measure. Once the optimal hyperparameters were found, each model was fit to the training dataset and then evaluated using root mean squared error and mean absolute error using the test dataset

Results

A comparison of the proposed stacked generalization model to the following models was then performed:

- Artificial Neural Network
- Support Vector Regressor
- Gradient Boost Regression Tree

When forecasting 24 hour ahead power output, regardless of seasonality, the ANN model and the SVR model exhibited very similar prediction behavior. However, in comparison to the ANN and SVR models the GBRT model seemed to have a relatively inverse relationship to the predicted values of the other two models. When ANN and SVR would predict a high value for a given observation the GBRT model would predict a lower value and vice versa. The stacked generalization model tended to produce predictions that closely resembled an average of the three predicted models. This can be expected considering that the final estimator, Ridge CV, for the stacked generalization model, is a form of linear regression. It is to be noted that all four models tended to fail to consistently predict peak power output. This is particularly apparent in the winter season.

The average performance of the individual models over all of the zones provided an RMSE of .119779 and a MAE of .065674. However, the stacked generalization model produced an average RMSE over all three zones of .11368 and a MAE of .06064.

Zone 1 Model Results				
	ANN	SVM	GBR	Stacked
R-Squared	0.798099	0.811395	0.79884	0.823859
RMSE	0.11625	0.112356	0.116036	0.10858
MAE	0.061427	0.059771	0.061593	0.056605

Zone 2 Model Results				
	ANN	SVM	GBR	Stacked
R-Squared	0.805423	0.804976	0.784248	0.820114
RMSE	0.121585	0.121725	0.12803	0.116905
MAE	0.066772	0.075127	0.068905	0.062794

Zone 3 Model Results				
	ANN	SVM	GBR	Stacked
R-Squared	0.812937	0.811517	0.810742	0.827376
RMSE	0.120289	0.120745	0.120993	0.115554
MAE	0.065663	0.064887	0.066919	0.062522

Model Average From All Zones				
	ANN	SVM	GBR	Stacked
R-Squared	0.805486	0.809296	0.797944	0.823783
RMSE	0.119375	0.118275	0.121686	0.11368
MAE	0.064621	0.066595	0.065806	0.06064

Summary				
	Zone 1	Zone 2	Zone 3	overall
R-Squared	0.823859	0.820114	0.827376	0.823783
RMSE	0.10858	0.116905	0.115554	0.11368
MAE	0.056605	0.062794	0.062522	0.06064

**Figure B - Evaluation Metrics for each model broken down by model and zone, by average performance over all three zones, and a summary of the stacked generalization model over all three zones.

As a result, of the benefit of being able to learn from the residuals of the other three models, a side by side comparison shows that the proposed model performed on average 5.4% better than each individual model when predicting power output 24 hours ahead. The stacked generalization model was able to weigh the benefits and weaknesses of each individual model to produce a more accurate output than each individual model.

Conclusions

In this paper a stacked generalization model containing an artificial neural network, support vector machine, and gradient boost regression tree as the first layer with the RidgeCV model used as the final estimator. Each model in the first layer model was trained separately for each zone. The final estimator was then trained on the predictions of the first layer models. In comparison to each individual model in layer one, the stacked generalization model performed on average 5.4% better. These experimental results demonstrate the improved model accuracy when the bias of each individual model can be mitigated by using a regressor in the second layer to produce a final prediction.

References

Carrera, Berny, and Kwanho Kim. "Comparison Analysis of Machine Learning Techniques for Photovoltaic Prediction Using Weather Sensor Data." Sensors, vol. 20, no. 11, 1 June 2020, p. 3129, doi:10.3390/s20113129.

Chen, Changsong, et al. "Online 24-h Solar Power Forecasting Based on Weather Type Classification Using Artificial Neural Network." Solar Energy, vol. 85, no. 11, Nov. 2011, pp. 2856-70, doi:10.1016/j.solener.2011.08.027.

Isaksson, Emil, and Mikael Karpe Conde. Solar Power Forecasting with Machine Learning Techniques. 2018, http://urn.kb.se/resolve?urn=urn:nbn:se:kth:diva-229065.

Matasci, Sara. "How solar panel cost and efficiency have changed over time." energysage - Smarter energy decisions, 16 July 2020, news.energysage.com/solar-panel-efficiency-cost-over-time/. Accessed 11 Dec. 2020.

"Nevada Question 6, Renewable Energy Standards Initiative (2020)." Ballotpedia, ballotpedia.org/Nevada_Question_6,_Renewable_Energy_Standards_Initiative_(2020). Accessed 11 Dec. 2020.

O'Leary, Daniel, and Joel Kubby. "Feature Selection and ANN Solar Power Prediction." Journal of Renewable Energy, vol. 2017, 8 Nov. 2017, pp. 1-7. Hindawi, doi:10.1155/2017/2437387. Accessed 11 Dec. 2020.

"Paris Agreement." Wikipedia, Wikimedia Foundation, en.wikipedia.org/wiki/Paris_Agreement#:~:text=This%20strategy%20involved%20energy%20and,20%25%20increase%20in%20energy%20efficiency.&text=The%20Paris%20deal%20is%20the%20world's%20first%20comprehensive%20climate%20agreement. Accessed 11 Dec. 2020.

Pedro, Hugo T.C., and Carlos F.M. Coimbra. "Assessment of Forecasting Techniques for Solar Power Production with No Exogenous Inputs." Solar Energy, vol. 86, no. 7, July 2012, pp. 2017-28, doi:10.1016/j.solener.2012.04.004.