Daniel Shannon
Proposal
CIS 600 Fundamentals of Data and Knowledge Mining
Syracuse University
May 9th, 2024

At Home Solar Forecasting: Personal Weather Station Regression-Fed Long Short Term Memory Solar Forecasting

https://github.com/radioxeth/photovoltaic-weather-station/tree/feature_extraction

Abstract
We aim to combine Personal Weather Station (PWS) and Residential Solar Power (RSP) data to forecast solar power generation using next-day local weather forecasts. Our approach is to train a Long Short-Term Memory (LTSM) neural network using 15-minute aggregated PWS and RSP data collected from January 2020 through March 2024. To reduce the dimensionality of the data, we will explore running polynomial regression (or random forest) using the Principal Components of the PWS and RSP data. The parameters of the regression will train the LSTM model. A sample of data is 24 hours of 15 minute interval data.

[*Results…*]

Introduction
Here we use three years of historical 5-minute PWS and 15-minute RSP data collected from a single family home in Denver, Colorado (**Figure 1)** to forecast next-day solar power output. We examine the use of Principal Component Analysis (PCA) and polynomial regression techniques to reduce the dimensionality of the 30-dimension dataset. (This might be a lot of work, and we stick with PCA).

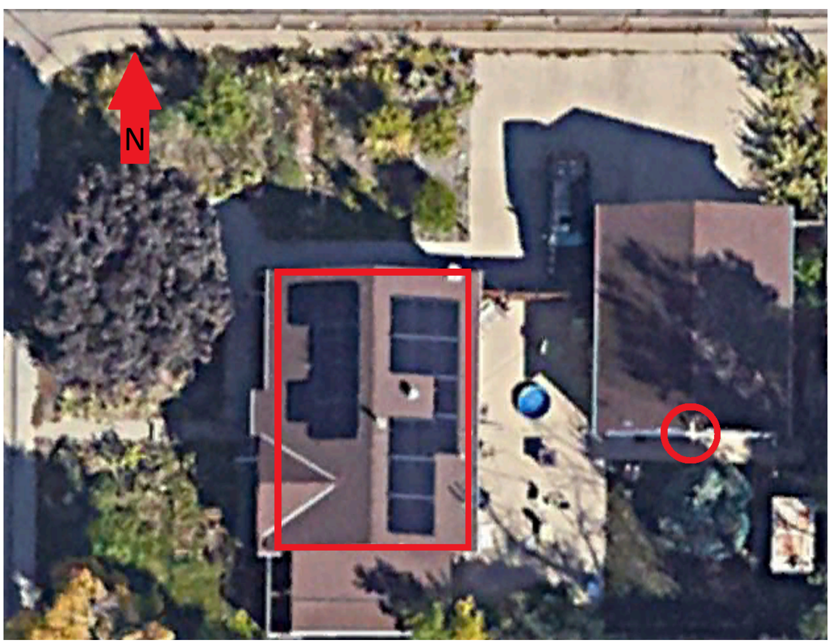


Figure 1. Residential solar panel array (box) and weather station (circle) subject location.

We then pass the parameters of the fit regression line to a LSTM Recurrent Neural Network, comparing the predictive results of the regression to the neural network. We measure the performance by comparing the cumulative predicted value vs the tenst value. The model will be trained on historical PWS data. To test, we can fit a regression line to the hourly forecast to get 15 minute interval data. We will measure performance as the forecasted cumulative power generated relative to the actual cumulative power generated.

Previous Work

[work in progress]

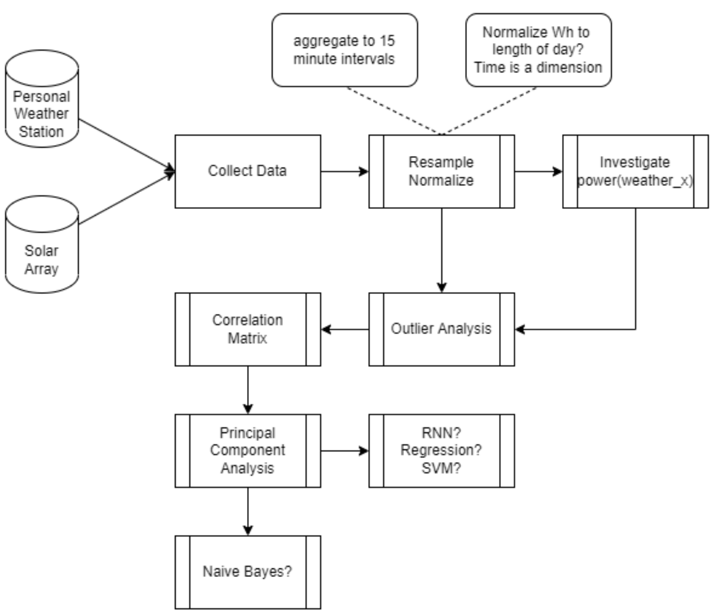
There are over 250,000 personal weather stations globally ("PWS Network Overview") and over 2.3 million residential solar installations in the US as of 2022 (Barbose et al.). As more individuals install RSP, there may arise the desire to forecast the performance of panels for the purpose of utility bill prediction and emergency preparedness in cases where the panels are connected to an on-site battery. Additionally, utilities need to understand the load of power generation relative to the capacity of the available infrastructure.

Forecasting Methods

Along with increasing residential solar panel installations, there is a growing interest in forecasting residential solar power generation. Classifiers such as random forests and decision tree models have been used to forecast solar and weather data. Random forests perform best out of statistical based methods when compared to linear regression, decision trees, and support vector machines, per Carrera et al.

Given the time series nature of solar power and weather data, several studies cite recurrent neural networks (RNN), and specifically Long Short Term Memory (LSTM) models (Bhagyanidhi et al.) (de Guia et al.) (Hossain and Mahmood).

I need to continue working on this section, but I have aggregated several papers in the works cited section...



Experiment Design

Design Overview

The overall experiment design consists of data collection, sampling, normalization, outlier analysis, dimension reduction, regression, training, and performance evaluation (**Figure 2**). Data collection is completed through RESTful APIs. Sampling, normalization, and outlier analysis level the sample to 15-minute intervals and reduce variance in the values. We explore the performance of regression and regression fed Long Short Term Memory neural networks for using weather based forecasts to predict solar power generation.

Figure 2. Experiment Design

Data Collection

We collect PWS and RSP data through their available Application Programming Interfaces (API). The PWS collects temperature, humidity, wind, humidity and pressure data and stores the data in 5-minute aggregated intervals. These parameters are available through the API aggregated as MAX, MIN, and MEAN values. UV intensity and solar irradiance are also available through the PWS API. Because the PWS used does not collect these data, they are returned as non values. The RSP data consists of power and energy, aggregated into 15-minute intervals.

Upon collecting the PWS and RSP data, we resample the PWS data from 5-minute intervals to 15-minute intervals using their respective aggregation functions - max, min, and mean. The end result of scraping and resampling the data is 31 dimensions, including local observation time as seen in **Figure 3**. After extracting and leveling the 15-minute data, we separate the data into 24 hour sections - resulting in 96 samples per section.

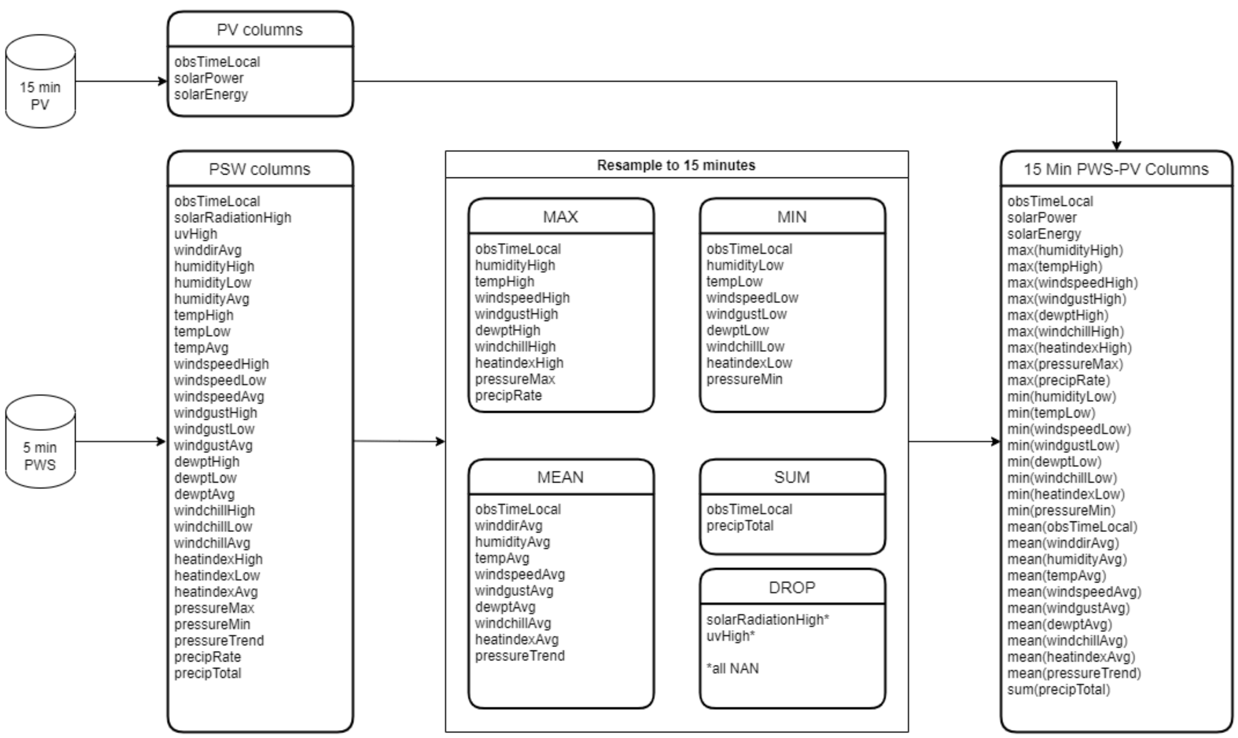


Figure 3. Extracted, translated, and leveled PWS and RSP data.

Dimension Reduction…

With 31 available dimensions, we can aim to reduce the number of dimensions either by manually selecting the dimensions or through Principal Component Analysis (PCA). With 3 years of 15-minute data, we have ~105,000 samples. Partitioning the samples into 24 hour sections over 3 years yields 1,095 samples. To approximate the number of dimensions needed to correctly fit the model to the data, we can use $2^d=n$, where d is the number of dimensions and n is the number of samples. With 1,095 samples, we can use 10 dimensions as a starting point.

10 dimensions can be found through PCA and we can examine the PCs by visually comparing the PC as a time sequence. We can achieve 6 dimensions by assessing the distinct dimensions collected by PWS, 1 dimension collected by RSP, and 3 dimensions calculated from measured PWS data, and observation time:

- **Observation time**
- PWS Measured
    - **Temperature**
    - **Humidity**
    - **Pressure**
    - **Wind speed**
    - **Wind direction**
    - **Precipitation**
- PWS Calculated
    - **Dew point**
    - **Head index**
    - **Wind chill**
- RSP
    - **Solar Power/Energy**

Regression and LSTM Model

We can define polynomial regression functions for each dimension to be used. Sinusoidal functions are an obvious choice, but may not capture the nuance of weather and solar data. We can feed the parameters of the regression to the LSTM model. If regression does not work, we can directly pass the 24 hour sections of weather data and classify them as the solar power generated for that day. (**Figure 4**.)

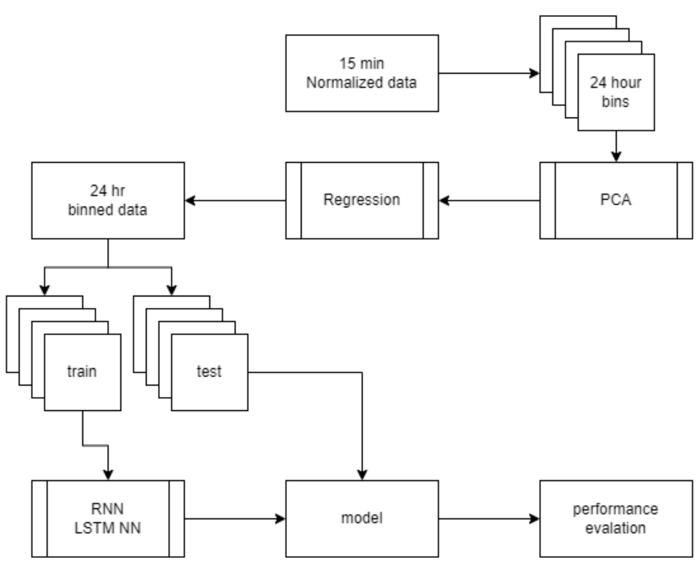


Figure 4.

Performance Evaluation
To evaluate the performance of the model, we can test using the data collected. We can split into 80/20, or find a better way to split the data. In order to test the model on real forecasts, we can take a 24 hour forecast and use regression to match 15-minute interval data. We pass the regressed forecasted weather into the model to predict the generated solar power.

Proposed Work
We need to begin testing the LSTM. We have begun training the model but we have yet to identify its performance. In addition to training and testing the model, we need to find a way to reliably collect next-day weather forecasts and sample it to 15 minute intervals. The roadblock I see is time, and so I may need to drop the regression and attempt to train on the raw data.

Works Cited

Barbose, Galen, et al. “Residential Solar-Adopter Income and Demographic Trends: 2022 Update.” *UC Open Access Publications*, 2022.

Bhagyanidhi, A., et al. “An Approach for Rainfall Prediction Using Long Short Term Memory Neural Network.” *2020 IEEE 5th International Conference on Computing Communication and Automation (ICCCA)*, 2020, https://ieeexplore-ieee-org.libezproxy2.syr.edu/document/9250809.

Carrera, Berny, et al. “PVHybNet: a hybrid framework for predicting photovoltaic power generation using both weather forecast and observation data.” *IET Renewable Power Generation*, vol. 14, no. 22, 2020, pp. 219-2201, https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/iet-rpg.2018.6174.

de Guia, J., et al. “Using Stacked Long Short Term Memory with Principal Component Analysis for Short Term Prediction of Solar Irradiance based on Weather Patterns.” *2020 IEEE REGION 10 CONFERENCE (TENCON)*, vol. 10.1109/TENCON50793.2020.9293719, 2020, https://ieeexplore.ieee.org/abstract/document/9293719.

Hossain, M., and H. Mahmood. “Short-Term Photovoltaic Power Forecasting Using an LSTM Neural Network and Synthetic Weather Forecast.” *IEEE Access*, vol. 8, 2020.

https://ieeexplore.ieee.org/abstract/document/9200614, 10.1109/ACCESS.2020.3024901.

Kim, Seul-Gi, et al. “A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning.” *MDPI Open Access Journals*, vol. 11, no. 5, 2019, https://www.mdpi.com/2071-1050/11/5/1501.

“PWS Network Overview.” *Weather Underground*, https://www.wunderground.com/pws/overview. Accessed 6 May 2024.

Razavi, S., et al. “From Load to Net Energy Forecasting: Short-Term Residential Forecasting for the Blend of Load and PV Behind the Meter.” *IEEE Access*, vol. 8, 2021, https://ieeexplore.ieee.org/abstract/document/9292948.