Analysis and Cost of Solar Power over Time

Proposal for Capstone Project 1

Springboard, DSCT

17Nov2017

Mark Sausville
saus@ieee.org

Analysis and Cost of Solar Power over Time

Capstone 1 Proposal

Overview

In this project, I will analyze a large dataset characterizing over 1 million solar power installations, targeting the cost over time of the roof mounted, residential installation to support the residential solar customer by predicting the future cost of a solar installation.

Structure

This project breaks down into the following steps:

1. Acquire the data.
2. Develop a mathematical model of the retail cost of a solar install to facilitate identification of the variables of interest within the data.
3. Identify the relevant data within the dataset and wrangle into a format that can be used for the following steps.
4. Perform a statistical learning analysis to find a function that expresses solar installation cost as a function of time and use that function to predict the cost of installation in the near term. Compare the performance of these various models and determine how they will be used to generate predictions. Also, study the effect of geographic location in the modeling.
5. Validate the methodology by backtesting on current and prior years.

Data Acquisition

The National Renewable Energy Laboratory (NREL) has a dataset containing relevant information concerning more than 1.02 million solar installations (see https://openpv.nrel.gov/search). This dataset contains 82 variables. Most of these variables are not pertinent to this analysis. There is little documentation of many of the fields, though some are self-describing.

In addition, NREL also maintains a public dataset that is somewhat smaller and better documented. This data has also been cleaned to the extent that certain fields have been canonicalized (i.e. multiple spellings for a single item have been removed).

Cost Model for Solar Installation

In market cycles we have come to expect that as the number of adopters grows, the cost of adoption drops as economies of scale come into play (producer cost/unit falls) and competition forces producers and installers to operate more efficiently.

In technology cycles, these effects are often exaggerated by innovation and discovery. Moore's Law is a well-known example: for many years now processor power/cost has roughly doubled every 18 months.

In solar energy, there is an analogous formulation, Swanson's law, i.e., the price of solar photovoltaic modules tends to drop 20 percent for every doubling of cumulative shipped volume. While not as dramatic a decrease as Moore's Law, drops in the cost of solar panels have been sufficient to have cut the price by 50% in the last decade.

For the customer, if the installation price is falling rapidly, waiting for the price to drop may be beneficial. On the other hand, if price is flat or increasing, buying now may be the best option. This study aims to educate that decision.

The cost of a solar installation has many components. The costliest single component has historically been the photovoltaic modules, but non-module costs currently amount to perhaps 2-3 times the module cost.

At a minimum, the following items must be considered in an analysis:

- Solar Panels
- Other hardware
  - Inverter
  - Racking
  - Cables, conduit, interconnection hardware, etc
- · Installation labor
- · Installer/Integrator costs and profit

Data Wrangling

I propose to isolate the variables above from the NREL dataset, removing irrelevant data.

In addition, some analysis must be performed to create a subset of the data with integrity, discarding conflicting data, but retaining all representative samples.

Statistical Analysis

Having selected the variables of interest, my method will be to statistically model each separately. Each cost segment in a solar installation has a price that will rise or fall independently over time. For example, the cost of solar panels has dropped dramatically over time though labor costs have stayed level or increased. Therefore, to predict the cost of the entire installation, each significant component of the cost must be modeled separately. The overall cost model is then the aggregate of the component models.

Since the data include final price, this can provide an immediate first order test of the aggregate model.

Backtesting

If the aggregate model is to be believed, it should be possible to predict the 2017 price from the data that precedes 2017 and so on.

If the predicted value closely corresponds with actual values, this will bolster the credibility of predictions for which there is not actual data.

Deliverables

The deliverables for this project will include the following:

- Final project report
- Project code
- Presentation slide deck