Screening Solar PV Adopters and Non-adopters: An Application of Machine Learning Methods

Changgui Dong, Renmin University of China, +86 10 82502304, changgui.dong@ruc.edu.cn
Benjamin Sigrin, National Renewable Energy Laboratory, +1 303 275 3244, Benjamin.Sigrin@nrel.gov

Overview

Though hardware costs for residential solar installations have dropped rapidly in recent years, non-hardware (soft) costs remain stubbornly high. Thus developing pathways for PV soft cost reduction is key to spurring more PV adoption and technology learning-by-doing. In particular, U.S. customer acquisition costs (CACs) represent a ripe area for improvement, with U.S. CACs around $0.62/W higher than equivalent German customers (Seel et al 2014). Improving efficiency of screening methods to differentiate potential PV adopters from non-adopters has the potential to save companies thousands of dollars per customers in CACs. Moreover, new survey data used in this paper provides important new insights to technology adoption literature, particularly, decompressing the customer decision process to reject adopting solar.

Most literature on PV demand only uses geographically aggregated data, usually at the zip code level (Bollinger and Gillingham, 2012; Davidson et al., 2014; Gillingham and Tsvetanov, 2015). Such methods cannot identify decision processes of individual PV adopters or non-adopters. Of studies that do employ adopter-level data, there are still some gaps. For example, Rai and Robinson (2013) used survey data to understand how much time it takes for PV adopters to make decision to adopt PV and the degree of peer influence; however, there was no data collected for non-adopters. In contrast, Zhai and Williams (2012) and Vasseur and Kemp (2015) collected some information from both adopters and non-adopters; nevertheless, they had focussed mainly on the perceived cost, maintenance and environmental concerns. These metrics are still very difficult for installers to use to screen potential customers without directly probing on each aspect.

Methods

This paper employs a novel survey dataset that was collected for the NREL-led DOE SEEDS project[1]. The dataset includes detailed information on three groups of households: those that have adopted PV, those who considered solar for their home but did not adopt, and those that have never considered PV for their home. For the purpose of this study, we focus on objective attributes of these correspondents that are easy-to-measure for PV installers, including variables such as summer electricity bill, household income and the squre footage of the house.

We use several machine-learning methods to classify people into different groups (adopters and non-adopters) based on those easy-to-measure attributes: recursive partitioning, random forest and generalized boosted models (GBM). These methods have been used extensively in medical diagnostic tests, health risk studies and other similar classification tasks, but is a new method for understanding new technology adoption decisions. An advantage of machine learning methods over logistic regressions is that they can consider a large number of variables simulataneously, including complex interaction effects across variables. The results from the former are also more intuitive, and generally more accurate.

Results

A typical decision tree result from recursive partitioning is presented in Figure 1 below, with the final nodes of interests (fraction of sample adopting) at the bottom. The stacked bar chart shows the division of people as either adopters (black) or non-adopters (white) and the size of the subset. Four attributes were used for classifying-- summer monthly electricity bill, the winter monthly electricity bill, household income level, and the square footage. As an example, node 13 contains no adopters, which is the set of respondents that have a summer bill greater less than $170/month, household income greater than $87,500, winter bill greater than $88/month, and a square footage greater than 3,400 ft^2. Anecdotally, such a property might constitute a second property that is infrequently used by the owner—who has little incentive to install solar.

[1] http://www.nrel.gov/SEEDS

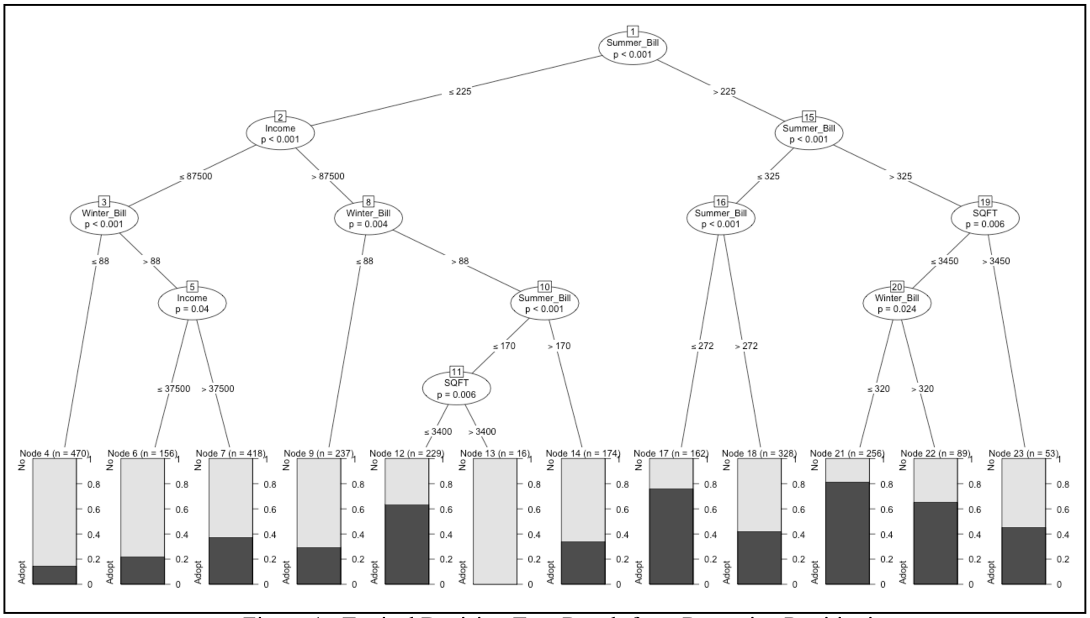

Figure 1. Typical Decision Tree Result from Recursive Partitioning

We can compare the accuracy of the classification results with actual data for cross-validation purposes, using logistic regression analysis as a benchmark. While the logistic regression model predicts the overall adoption status (Yes or No) correctly 67% of the time, it only correctly predicts 41% of adopters . In contrast, the Recursive Partitioning model was more accurate along both measures, 71% and 49% respectively. Furthermore, the Random Forest model increased the correct rate for adopters to 72% (more than 300 adopters being correctly classified), and GBM elevated the correct rate to 85%. In addition to these binary results, the model produces detailed probabilistic results for each customer to become an adopter; in constrast to results from a typical logistic regression model, these probabilities are not necessarily monotocially increasing or decreasing with any single attribute of customers.

Our survey results also covered why people choose not to adopt solar PV. So far, the top reasons include: 1) upfront costs are too high; 2) the electricity bill is too small; 3) the roof is not suitable to install PV panels; and 4) might move before the investment pays off. These reasons could further inform our machine learning methods in variable selection and facilitate interpreting the variable importance results from emsenble methods such as Random Forest.

Conclusions

To further reduce soft cost of solar PV, new methods must be developed to improve efficiency in identifying prospective customers. The industry has recently seen a number lead-generation market entrants hoping to capitalize on this opportunity. The application of machine learning methods has the potential to help the industry lower marketing costs based on easy-to-measure customer attributes. When combined with reasons why people choose not to adopt solar PV, the results from this study could be very useful in screening adopters and non-adopters.

References

Bollinger, Bryan, and Kenneth Gillingham. 2012. “Peer Effects in the Diffusion of Solar Photovoltaic Panels.” Marketing Science 31 (6): 900–912.

Davidson, Carolyn, Easan Drury, Anthony Lopez, Ryan Elmore, and Robert Margolis. 2014. “Modeling Photovoltaic Diffusion: An Analysis of Geospatial Datasets.” Environmental Research Letters 9 (7): 074009.

Gillingham K., Tsvetanov T., 2015. Hurdles and Steps: Estimating Demand for Solar Photovoltaics, available at: http://environment.yale.edu/gillingham/GillinghamTsvetanov_SolarDemandCT.pdf.

Rai, Varun, and Scott A. Robinson. 2013. “Effective Information Channels for Reducing Costs of Environmentally-Friendly Technologies: Evidence from Residential PV Markets.” Environmental Research Letters 8 (1): 014044.

Seel, Joachim, Galen L. Barbose, and Ryan H. Wiser. 2014. “An Analysis of Residential PV System Price Differences between the United States and Germany.” Energy Policy 69 (June): 216–26.

Vasseur, Véronique, and René Kemp. 2015. “The Adoption of PV in the Netherlands: A Statistical Analysis of Adoption Factors.” Renewable and Sustainable Energy Reviews 41 (January): 483–94.

Zhai, Pei, and Eric D. Williams. 2012. “Analyzing Consumer Acceptance of Photovoltaics (PV) Using Fuzzy Logic Model.” Renewable Energy 41 (May): 350–57.