Towards Data-Driven Energy Consumption Forecasting of Multi-Family Residential Buildings: Feature Selection via The Lasso

R. K. Jain[1], T. Damoulas[1,2] and C. E. Kontokosta[1]

[1]Center for Urban Science + Progress, New York University, 1 MetroTech Center, 19th Floor, Brooklyn, NY 11201; email: {rishee.jain, damouslas, ckontokosta}@nyu.edu
[2]Polytechnic School of Engineering, New York University, 2 MetroTech Center, 10th Floor, Brooklyn, NY 11201

ABSTRACT

Buildings constitute a large portion of energy consumption in the United States. Accurate forecasting of building energy consumption is integral to the implementation of energy efficiency initiatives and intermittent renewable energy supplies. The availability of high resolution energy consumption data has allowed researchers to utilize machine learning techniques that forego domain specific knowledge (e.g., building construction materials, geometric properties) to forecast energy consumption. While there is a growing body of literature surrounding the use of machine learning to forecast building energy consumption, previous research has yet to explore the use of feature selection to determine the most important subset of variables and produce interpretable predictive models. In this paper, we explore the use of Lasso, a shrinkage and selection method for linear regression that estimates sparse coefficients, to select the most important feature subset of a residential energy forecasting model. We evaluate the selected subset on an empirical data set from a multi-family residential building in New York City and compare the results to previous forecasting models without feature selection. Results of this work has implications on the data acquisition and sensing systems required to yield accurate predictions of residential energy consumption.

INTRODUCTION

Buildings account for over 40% of energy consumption and GHG emissions in the United States (U. S. Department of Energy, 2011). Thus, understanding and forecasting energy consumption in buildings is paramount to improving energy efficiency and overall sustainability.

Traditionally, researchers have relied on engineering software packages (e.g., EnergyPlus) to model energy consumption in buildings. While such approaches are valuable, they requires a significant amount of information and data regarding the structural, geometric and material properties of a building. Obtaining and validating such data is a hindrance to wide spread energy forecasting. As an alternative, researchers have begun exploring “data-driven” approaches to model and forecast energy consumption. Such data-driven approaches forgo the need for exhaustive input data on the physical properties of a building. Smart meters are utilized to capture high resolution energy consumption data that is fed into a machine learning algorithm to infer the complex relationships between energy consumption and other influencing variables such as temperature, time of day and occupancy.

While there is a growing body of literature regarding data-driven forecasting techniques, there has been limited research applying feature selection to determine the most important subset of variables and produce more interpretable predictive models. *In this paper, we explore the use of Lasso, a shrinkage and selection method for linear regression that that estimates sparse coefficients, to select the most important feature subset of a multi-family residential energy forecasting model.* Additionally, we compare the results to previous applications of data-driven forecasting models to determine if comparable accuracy can be achieved without the use of any additional external variables (e.g., temperature, occupancy).

BACKGROUND

There has been an extensive amount of work that examines data-driven energy forecasting for commercial buildings and to a lesser extent residential buildings. The concept of data-driven energy forecasting was introduced as part of the Great Energy Predictor Shootout (Kreider and Haberl, 1994). Participants were asked to predict the electricity consumption of a commercial building based on historical consumption and environmental data. Several studies extended this work in the commercial building (Dong et al., 2005; Karatasou et al., 2006; Li et al., 2009; Wong et al., 2010; Yang et al., 2005) and residential building (Edwards et al., 2012) sectors.

Specifically, Edwards et al. (2012) applied several data-driven forecasting techniques to predict the energy consumption of three residential single-family homes. While Edwards et al. (2012) was able to achieve high levels of accuracy, their predictive models relied on the use of input data from over 140 sensors placed throughout the single-family homes. Installing such sensing devices on a large scale would be both intrusive and cost-prohibitive and in turn limits the applicability of such models to applications of wide-spread energy consumption forecasting. Limiting the required inputs in a forecasting model to only energy consumption data would not necessitate the installation of sensing devices beyond energy smart meters.

Previous work has explored the use of feature selection methods in the domain of building energy consumption forecasting (Kolter and Ferreira, 2011; Zhao and Magoules, 2012). Kolter and Ferreira (2011) employed a simple greedy forward feature selection procedure to determine the best features for predication of monthly energy consumption of 6,500 buildings in Cambridge, MA. Application of feature selection reduced the number of features from 35 to 9 and yielded a more interpretable predictive model. Zhao and Magoules (2012) extended feature selection to high-fidelity and more temporally granular energy data through the application of statistical learning methods (i.e., Correlation Coefficient (CC), Regression, Gradient guided feature selection (RGS)). Application of such statistical learning methods allowed for prediction accuracy to be maintained while reducing the model features from 23 to 6. While feature selection successfully reduced the number of required features, the resultant model still required data from additional sensors that may be intrusive and difficult to implement on a large number of buildings (e.g., outdoor air density, water main temperature, light total heat gain). In this paper, we extend previous work in feature selection to explore how energy forecasting models can be developed with a limited amount of data from additional sensors.

Additionally, multi-family residential buildings introduce an added level of complexity over single-family homes. Multi-family residential buildings by definition are comprised of numerous independent apartment units. Each individual unit has its own independent characteristics and subsequent energy consumption patterns. Thus, multi-family residential buildings can be modeled on two different spatial scales – whole building and unit-by-unit. In order to further understand how such spatial scales impact feature selection and forecasting models, we deepen our analysis and exploration to encompass predictions on multiple spatial scales in this paper.

METHODOLOGY

We aim to apply feature selection for forecasting the energy consumption of a multi-family residential building. Specifically, we aim to apply the Lasso, a shrinkage and feature selection method for linear regression that estimates sparse coefficients, to explore how the features of a forecasting model can be reduced to increase applicability for wide-spread energy forecasting. We develop a forecasting model using the Sckit-learn module (Pedregosa et al., 2011) in the Python development language and tested several scenarios using an empirical dataset collected from a multi-family residential building in New York City.

Model features and coefficient of variation (CV) performance metric. Based on previous work (Edwards et al., 2012), we define the following features as the baseline prior to employing the Lasso feature selection method (Table 1).

Table 1. Model features before feature selection

Feature	Description
T(t)	Current outdoor temperature
S	Indicator variable denoting weekend/holiday
Sh	Sine of the current hour
Ch	Cosine of the current hour
y(t-1)	Energy consumption at time (t-1)
y(t-2)	Energy consumption at time (t-2)
y(t-3)	Energy consumption at time (t-3)
y(t-4)	Energy consumption at time (t-4)
y(t-5)	Energy consumption at time (t-5)

The time step will vary depending on the temporal base of each scenario. For example, for the hourly interval model the energy consumption *y(t-1)* indicates the energy consumption for the previous hour.

In order to assess the performance of our model and compare results with previous work, we utilize the coefficient of variation (CV) as our performance metric. This is consistent with several previous studies that utilized CV (Edwards et al., 2012; Karatasou et al., 2006; Kreider and Haberl, 1994) as the basis for assessing performance. The CV metric is defined as follows:

$$CV = \frac{\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(y_i - \widehat{y_i})^2}}{\bar{y}} * 100$$

where: $\widehat{y_i}$ is the predicted value, y_i is the observed value, $\bar{y}$ is the mean of the observed values and N is the total number of observations.

Mechanics of Lasso feature selection. We examine the use of a prominent shrinkage and selection method for linear regression models called "The Lasso" (Tibshirani, 1996). The technique has been successfully applied to many domains since its introduction as it enjoys the benefits of being an explicitly sparse method (some regression coefficients set to 0) that leads to interpretable models. In addition, as we will show in this paper for the problem of predicting energy consumption, it can be competitive to more complex (kernel-based) models like Support Vector Regression (SVR) while utilizing a much smaller subset of variables and having lower computational complexity.

The Lasso is a linear regression method that minimizes the usual squared error loss plus an L1 penalty on the regression coefficients. The minimization of the overall loss function is given by:

$$\hat{\beta} \leftarrow \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{n=1}^{N} \left(y_n - b - \sum_{d=1}^{D} \beta_d x_{dn} \right)^2 \right\} + \lambda ||\beta||_1$$

where $\{x_n, y_n\}_{n=1}^{N}$ is the training data, b the intercept, and the Lagrange multiplier λ balances the tradeoff between the Squared Error loss and the L1 penalty $||\beta||_1$ on the regression coefficients.

The effect of the L1 norm is to impose a competition between regression coefficients and effectively shrink some of them to 0 hence producing a sparse model that explains the data with as little variables as possible. In contrast with kernel-based SVR that learns a sparse model in terms of data samples (support vectors), the Lasso learns a sparse model in terms of features.

Model optimization. The Lagrange multiplier (λ) is typically inferred from cross-validation. We performed this optimization using the *k-fold* cross validation method to avoid over fitting our model. This method splits the training set into *k* sequential subsets of near-equal size, holding out each in turn as a validation set, while using the other *k-1* subsets to train the model. In our application of this method, we set *k* equal to 5 folds.

Application on empirical dataset. The empirical dataset to be utilized in this paper consists of electrical consumption data for 21 units in a six story multi-family residential building (Watt Hall) on the Columbia University campus in New York City. Electrical consumption is captured in temporal intervals of 10 minutes for each individual unit from August 27, 2012 to December 19, 2012 (84 days). The energy consumption monitoring and data acquisition system is detailed in Gulbinas et al. (2013). Electrical consumption data was supplemented with hourly outdoor temperature data downloaded from the National Oceanic and Atmospheric Association (NOAA) Central Park Weather Station. Because the building was occupied by undergraduate students, both university-wide holidays and weekends were encoded into the model's *s* indicator variable. Watt Hall is a typical pre-war multi-family residential building in New York City with high ceiling and thick plaster walls. All units in the building have access to natural light either via a courtyard or the street and are comprised of a kitchen, bathroom, living area and bedroom area.

Scenario testing. We tested our feature selection model on the scenarios described in Table 2. The temporal granularity reflects the temporal scale in which data is being inputted into the forecasting model and predictions are made. For aggregated building scenarios (A, C), we aggregated the data by summing all 21 units and then trained the model on the aggregated values. For testing the individual unit scenarios (B, D), we developed and trained an independent model for each individual unit. In order to understand the impact feature selection has on the performance of our prediction model, we compare the results with previous work that applies Support Vector Regression (SVR) on the same empirical dataset (Jain et al., 2013; Jain et al., under review).

Table 2. Temporal and Spatial Scenarios

Scenario	Temporal Scale	Spatial Scale
A	Hourly	Building
B	Hourly	Unit
C	10 mins	Building
D	10 mins	Unit

RESULTS

Overall results. Our results indicate that the Lasso is able to discover a very sparse representation that effectively utilizes a very small subset of the variables, offering intuition about their predictive capabilities and also competing performance with SVR. Across all scenarios, application of the Lasso illustrated that the most dominant feature for our energy prediction model was the previous time step's energy consumption *y(t-1)*. This finding held constant across both hourly and 10 min temporal intervals as well as both aggregated building and individual unit spatial granularity. Details on the performance of specific scenarios are outlined in the subsequent sections.

The full Lasso path (illustrated in Figure 1) shows the values of the regression coefficients for increasing values of λ as the penalty contribution increases. The dotted line indicates the Lagrange multiplier (λ) chosen by *k-folds* cross-validation. The only line present on the plot is that corresponding to the *y(t-1)* feature. Thus, it can be inferred that other influencing features were found to be non-influential very early in the Lasso feature selection process.

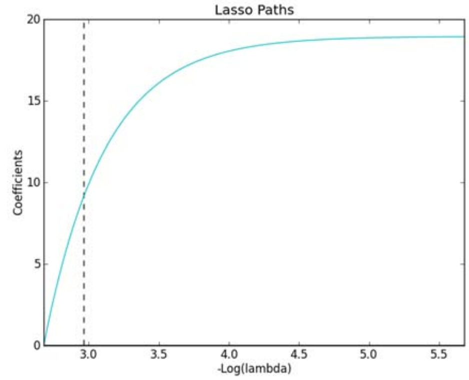

Figure 1. Lasso feature selection path for scenario C

Hourly scenarios (A, B). Results for the hourly scenarios (A, B) indicate that the Lasso performs on par with an SVR model. Results for predicting the aggregated building consumption (scenario A) are provided in Figure 2 with CV values found to be consistent between the Lasso (CV=12.03) and the SVR methods (CV=11.31). Similar results were found when the model was applied to predicting individual unit consumption (scenario B). The results for a sample unit (4H) are provided in Figure 3. The CV values were found to be consistent across both methods with the Lasso (CV=97.39) fairing slightly worse than the SVR model (CV=88.71) for this particular unit.

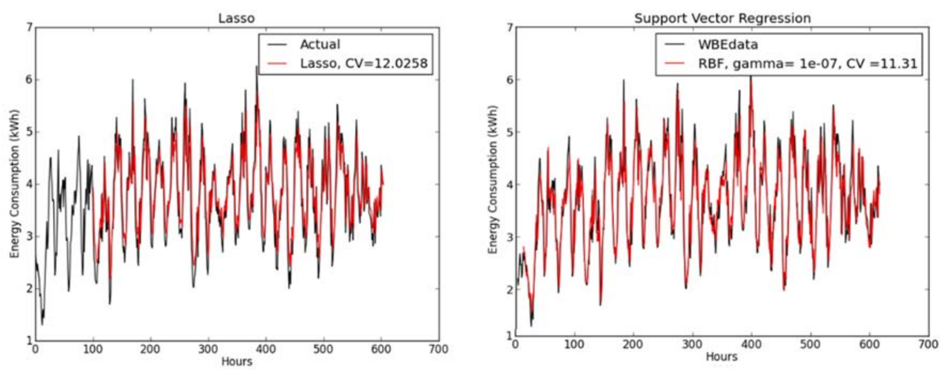

Figure 2. Results of Lasso and SVR methods for Scenario A

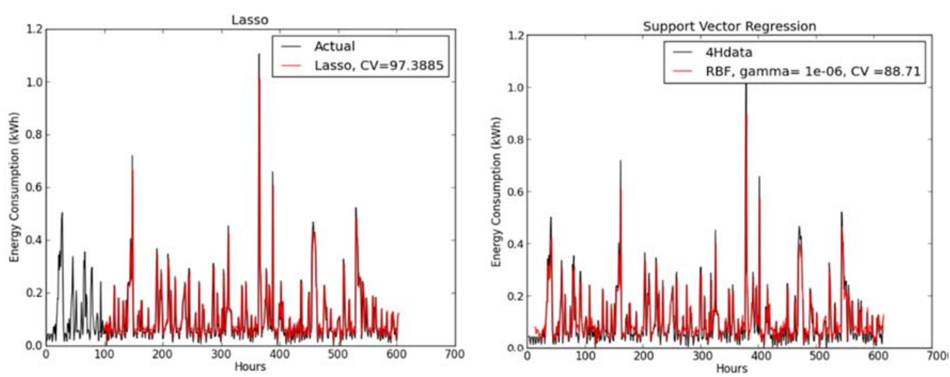

Figure 3. Results of Lasso and SVR methods for Scenario B (sample unit 4H)

10 min scenarios (C, D). Results for the aggregated building prediction at 10 min temporal intervals (scenario C) are provided in Figure 4. Results were consistent with the hourly results with the Lasso (CV=14.88) in this case performing marginally worse than the SVR model (CV=10.66). However, for the individual unit predictions (scenario D), the Lasso surprisingly outperformed the SVR model despite utilizing the *y(t-1)* as the dominant predictive feature. The results for a sample unit (4H) are provided in Figure 5. In the case of unit 4H, the Lasso (CV=86.18) is shown to perform substantially better than the SVR model (CV=202.29).

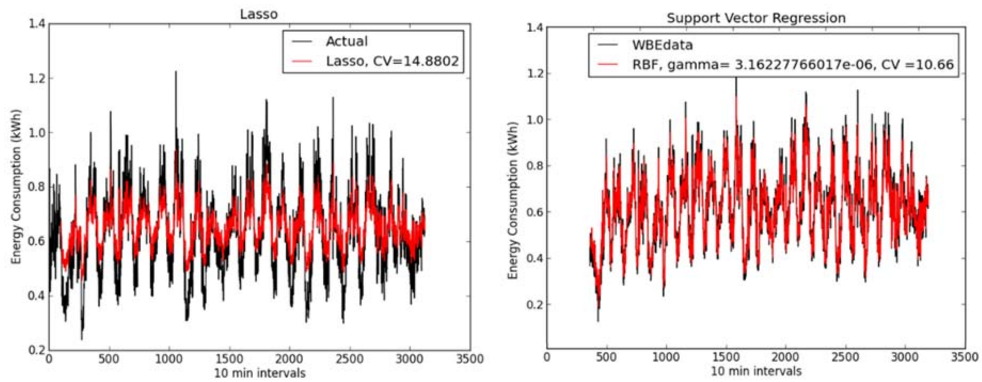

Figure 4. Results of Lasso and SVR methods for Scenario C

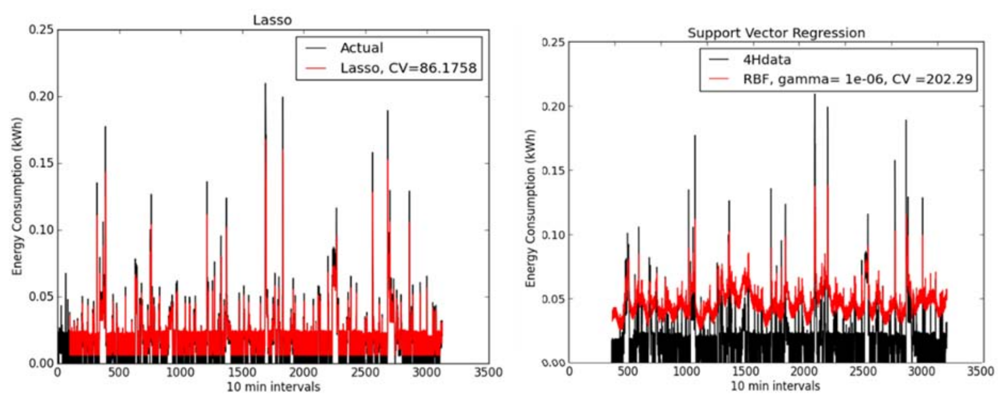

Figure 5. Results of Lasso and SVR methods for Scenario D (sample unit 4H)

CONCLUSION

The results of this work provide insight into what features are necessary to achieve accurate predictions of building energy consumption. Through application of the Lasso, we were able to demonstrate that accurate predictions can be achieved without data from external sensors (e.g., temperature, occupancy, building size). Moreover, we found that the Lasso with only the previous time step's consumption *y(t-1)* as input surprisingly outperforms an SVR model in certain temporally and spatially granular cases (scenario D). This work represents a first step in understanding how accurate energy forecasting can be achieved with minimal input data and has implications for the deployment of energy data acquisition and sensing systems. Future research is necessary to further explore and refine the application of feature selection methods, such as the Lasso, on additional empirical data sets.

Understanding, characterizing and forecasting energy consumption could provide the missing link necessary to implement advanced building energy efficiency measures.

REFERENCES

Dong B, Cao C, Lee SE. (2005). Applying support vector machines to predict building energy consumption in tropical region. *Energy and Buildings*. 37:545–53.

Edwards RE, New J, Parker LE. (2012). Predicting future hourly residential electrical consumption: A machine learning case study. *Energy and Buildings*. 49:591–603.

Gulbinas, R., Jain, R. K., Taylor, J. E., Peschiera, G., & Golparvar-Fard, M. (2013). Network Eco-Informatics: Development of a Social Eco-Feedback System to Drive Energy Efficiency in Residential Buildings. *Journal of Computing in Civil Engineering*.

Jain, R., Smith, K., Culligan, P. and Taylor, J. (2013). Exploring Energy Consumption Forecasting for Multi-Family Residential Buildings Using Support Vector Regression. *Proceedings of the 8th Conference on Sustainable Development of Energy, Water and Environment Systems*.

Jain, R., Smith, K., Culligan, P. and Taylor, J. (under review). Forecasting Energy Consumption of Multi-Family Residential Buildings Using Machine Learning: Investigating the Impact of Temporal and Spatial Monitoring Granularity on Accuracy Performance.

Karatasou S, Santamouris M, Geros V. (2006). Modeling and predicting building's energy use with artificial neural networks: Methods and results. *Energy and Buildings*. 38:949–58.

Kolter J, Ferreira J. (2011). A large-scale study on predicting and contextualizing building energy usage. *Proceedings of the 25th Conference on Artificial Intelligence*. 1349–56.

Kreider JF, Haberl JS. Predicting hourly building energy usage. *ASHRAE Journal*. 36.

Li Q, Meng Q, Cai J, Yoshino H, Mochida A. (2009). Predicting hourly cooling load in the building: A comparison of support vector machine and different artificial neural networks. *Energy Conversion and Management*. 50:90–6.

Pedregosa F, Varoquaux G., Gramfort, A… (2011). Scikit-learn: Machine learning in Python. *JMLR*. 2825-2830

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*. 267-288.

U.S. Department of Energy. (2011). Buildings Energy Data Book.

Wong SL, Wan KKW, Lam TNT. (2010). Artificial neural networks for energy analysis of office buildings with daylighting. *Applied Energy*. 87:551–7.

Yang J, Rivard H, Zmeureanu R. (2005). On-line building energy prediction using adaptive artificial neural networks. *Energy and Buildings*. 37:1250–9.

Zhao H, Magoulès F. (2012). Feature Selection for Predicting Building Energy Consumption Based on Statistical Learning Method. *Journal of Algorithms & Computational* Technology. 6:59–78.