World Academy of Science, Engineering and Technology
International Journal of Energy and Power Engineering
Vol:11, No:2, 2017

Development of Prediction Models of Day-Ahead Hourly Building Electricity Consumption and Peak Power Demand Using the Machine Learning Method

Dalin Si, Azizan Aziz, Bertrand Lasternas

***Abstract*—**To encourage building owners to purchase electricity at the wholesale market and reduce building peak demand, this study aims to develop models that predict day-ahead hourly electricity consumption and demand using artificial neural network (ANN) and support vector machine (SVM). All prediction models are built in Python, with tool Scikit-learn and Pybrain. The input data for both consumption and demand prediction are time stamp, outdoor dry bulb temperature, relative humidity, air handling unit (AHU), supply air temperature and solar radiation. Solar radiation, which is unavailable a day-ahead, is predicted at first, and then this estimation is used as an input to predict consumption and demand. Models to predict consumption and demand are trained in both SVM and ANN, and depend on cooling or heating, weekdays or weekends. The results show that ANN is the better option for both consumption and demand prediction. It can achieve 15.50% to 20.03% coefficient of variance of root mean square error (CVRMSE) for consumption prediction and 22.89% to 32.42% CVRMSE for demand prediction, respectively. To conclude, the presented models have potential to help building owners to purchase electricity at the wholesale market, but they are not robust when used in demand response control.

***Keywords*—**Building energy prediction, data mining, demand response, electricity market.

I. Introduction

INCREASINGLY intensive energy consumption is a global issue which causes environmental degradation, resource crisis, and ozone depletion. Energy consumed by buildings comprises a large portion of the total consumption. According to research from the International Energy Agency, buildings account for 40% of total primary energy consumption [1]. In addition, the energy consumed varies significantly during the course of one year, especially for buildings. For example, in the course of a summer's day, the electricity consumed during the on-peak period is more than 50% of the total daily consumption [2] and can result in power grid stress.

Specifying to electricity usage in building, electricity shortage and waste are mainly caused by the mismatch between the electricity supply and demand. Therefore, the power generation industry has to capture that varying relationship between demand and supply all the time. Owing to this continually changing relationship between the supply and demand of electricity, the cost of electricity consumption to consumers varies continuously [3]. To solve this problem, a dynamic pricing schedule was proposed, in which prices are more expensive during on-peak periods and cheaper during off-peak periods. This dynamic pricing schedule can achieve an economical balance between the supply and demand, and also provide an incentive for consumers to change their electricity usage pattern to ease grid stress. Now, many of the notable markets, such as NYISO, PJM, Interconnection, ERCOT in the United States, use dynamic pricing schedules on the wholesale market.

Technically, the dynamic pricing schedule mainly exists at the wholesale market. In general, the US electricity market consists of a wholesale market and a retail market. A wholesale market exists when competing electricity suppliers offer their electricity to utilities. Then, utilities sell the electricity they bought from the wholesale market to consumers on the retail market. More specifically, the wholesale market consists of two dynamic pricing markets, the day-ahead market and the real time market. At the day-ahead market, utilities forecast their consumers' demand and purchase the same amount of electricity. The difference between forecasted consumption and actual consumption is paid at the real time price on the real time market [4].

From the consumers' point of view [5], rather than buying electricity from utilities at the retail market, buying electricity at wholesale market provides them both opportunities and risks. If the consumer can predict their hourly demand and consumption a day-ahead, they only need to pay a relatively cheap price at the day-ahead wholesale market, rather than buy at an expensive price from utilities on the retail market. However, any forecast error and other uncertainty could expose the consumer to risk in its wholesale market transactions. Therefore, a tool is needed to precisely predict demand and consumption a day-ahead to help consumers avoid those uncertain risks and even get financial benefit in their electricity purchases.

Besides helping consumers get benefits in the wholesale market, demand prediction can also help customers to change their usual usage pattern to reduce demand during the on-peak period. This kind of control is referred to as the demand response (DR) control. In terms of different DR control strategies, in common sense, there are three basic kinds of strategies that could be achieved. The first type is demand reduction, reducing the electricity load during the on-peak period when the price is high and maintaining the load pattern during the off-peak period. Even though this type could temporarily sacrifice occupant comfort, it has the broadest

D. S. Author is with the Intelligent Workplace, Carnegie Mellon University, Pittsburgh, PA 15213 USA (phone: 412-961-4626; e-mail: sdl0619@hotmail.com).

application. As an example, the thermal set-point of HVAC system could be temporarily changed during the on-peak period [6]. Second, the DR controller can reschedule the operation pattern, and shift the dispensable load from the on-peak period to off-peak period. For instance, some household activities, such as dishwasher, laundry dryer, or pool pump can be shifted from noon to night. However, buildings this type fitted are only residential buildings or hotels. It cannot be the case for commercial or industrial buildings, because most of the load cannot be shifted and fixed with schedule. The last type is peak supplement; the DR control can do this by generating electricity (PV system, wind turbine) or discharging electricity (ice or battery) on site. As an application example, the first control strategy will be applied on the proposed model in this research.

The proposed model aims to predict building hourly consumption and day-ahead demand. Generally, there are four methods for electricity demand and consumption forecasting [7]: engineering method, statistical method, ANN and SVM. These methods can be grouped into two methods, engineering methods and machine learning methods. Engineering methods use physical principles to calculate the energy performance of buildings. Hundreds of software programs have been developed for evaluating energy performance and analyzing building energy efficiency, such as eQUEST, EnegryPlus, ESP-r. However, engineering methods require too much information, and modeling and further calibration, which is time consuming. In contrast, as a quite novel approach, machine learning methods can acquire the relevant inputs which could most affect the output and find the relationship between them. It is an interdisciplinary field, which includes the content of mathematics, statistics, and data visualization. In the past decades, a lot of researches about demand or consumption forecasting of buildings have been investigated. Dong Bing's paper [8] mainly focuses on estimating monthly energy usage using SVM techniques, and his result shows that SVM has a good performance in prediction. Yokoyama [9] *et al.* proposed the "Model Trimming Method," a global optimization method to optimized ANN, and also showed a good performance, but only focused on the cooling demand. However, if in view of the application on electricity purchase or DR, few papers consider the feasibility issue. Therefore, this research develops a feasible approach to predict hourly building consumption and demand day-ahead. Models for day-ahead hourly demand and consumption by SVM and ANN will be developed, and a DR algorithm will also be proposed as an application.

II. Methodology

A. Background Information

The target building in this research is the Center for Sustainable Landscape (CSL). It is a three floors building, located at Schenley Drive, Pittsburgh. It is the multi-functional building that contains an office, art gallery, and classroom. The motivation of this research is mainly based on the dynamic electricity pricing, and determines the electricity buying strategy and DR algorithm. Many utilities or companies in the US offer dynamic electricity pricing [10]-[12], and these services are based on the day-ahead or real-time wholesale price. The day-ahead or the real-time markets are the wholesale electricity markets which are operated by the Regional Transmission Organization (RTO) or Independent System Operator (ISO). The day-ahead price (price changes hourly) is decided one day in advance according to the day ahead hourly biding price from low to high. The real-time market sets the electricity prices at buses or nodes based on real conditions of demand and supply every five minutes. The working principle of this market is that, for example [13]: the day-ahead price (DAP) is $20, the schedule demand (predicted next day demand) is 100MW, but the next day actual demand is 105MW and the real-time price (RTP) is $23, the total charge should be: 100*20+(105-100)*23=$2,115.

B. Research Outline

Based on the information of the target building and electricity market, the research outline is proposed and its schematic outline is shown in Fig. 1. As shown in different colors, this research has three main parts generally. All algorithms used in this research are implemented in Python, with assisted by the packages, Scikit-learn [14], [15] and Pybrain [16].

First, all the processes in the red background are the energy prediction part. In this part, one year CSL relevant historical data, from 2015-01-01 to 2015-12-31, with 1 hour intervals is collected as a training dataset. (The demand is collected from the peak value within each hour.) The features in the training data-set are: time stamp, outdoor dry bulb temperature, outdoor relative humidity, AHU supply air set-point, solar radiation (five features), electricity consumption and hourly peak demand (two outputs). These features were selected based on the common points of literature review. The first step in this section is a primary training. All data used in the primary training are historical data, which can be easily obtained from the CSL weather station. Because in reality, some features cannot be obtained a day-ahead (such as solar radiation), the models trained in the primary training are not the final results. The aims of this primary training are: 1. Find a better model between SVM and ANN; 2. As baseline models to compare with the final one.

In practical, those historic data in the training dataset should also be available a day ahead to predict next day peak demand and consumption. However, solar radiation is the only unavailable day-ahead feature. Although we can easily get historical hourly solar radiation from weather station records, there is no hourly solar radiation provided in weather forecasting. Therefore, we have to predict the next day solar radiation first, and then using predicted solar radiation as input to estimate demand and consumption. That is what we plan to do in the second part, which is for solar radiation and is in the yellow background in Fig. 1. For next day solar radiation prediction, two methods are examined, they are the engineering model (Zhang-Huang solar model [17]) and the boosted regression tree (BRT) model, the better one will be chosen. The inputs for solar radiation prediction, such as forecasting of wind speed, sky cover and wind direction are acquired from weather forecasting API ForecastIO [18]. Then, returning to the energy prediction part, the final training will be implemented. Predicted solar radiation, will substitute the historical data, and be re-organized with other features mentioned in the primary training as the new dataset. The new dataset will be trained by the better method (ANN or SVM) obtained from the primary training. After that, the results from this new dataset will be compared with the results from primary training. The comparison results will manifest whether the predicted solar radiation is competent to substitute the historical one to predict demand or consumption.

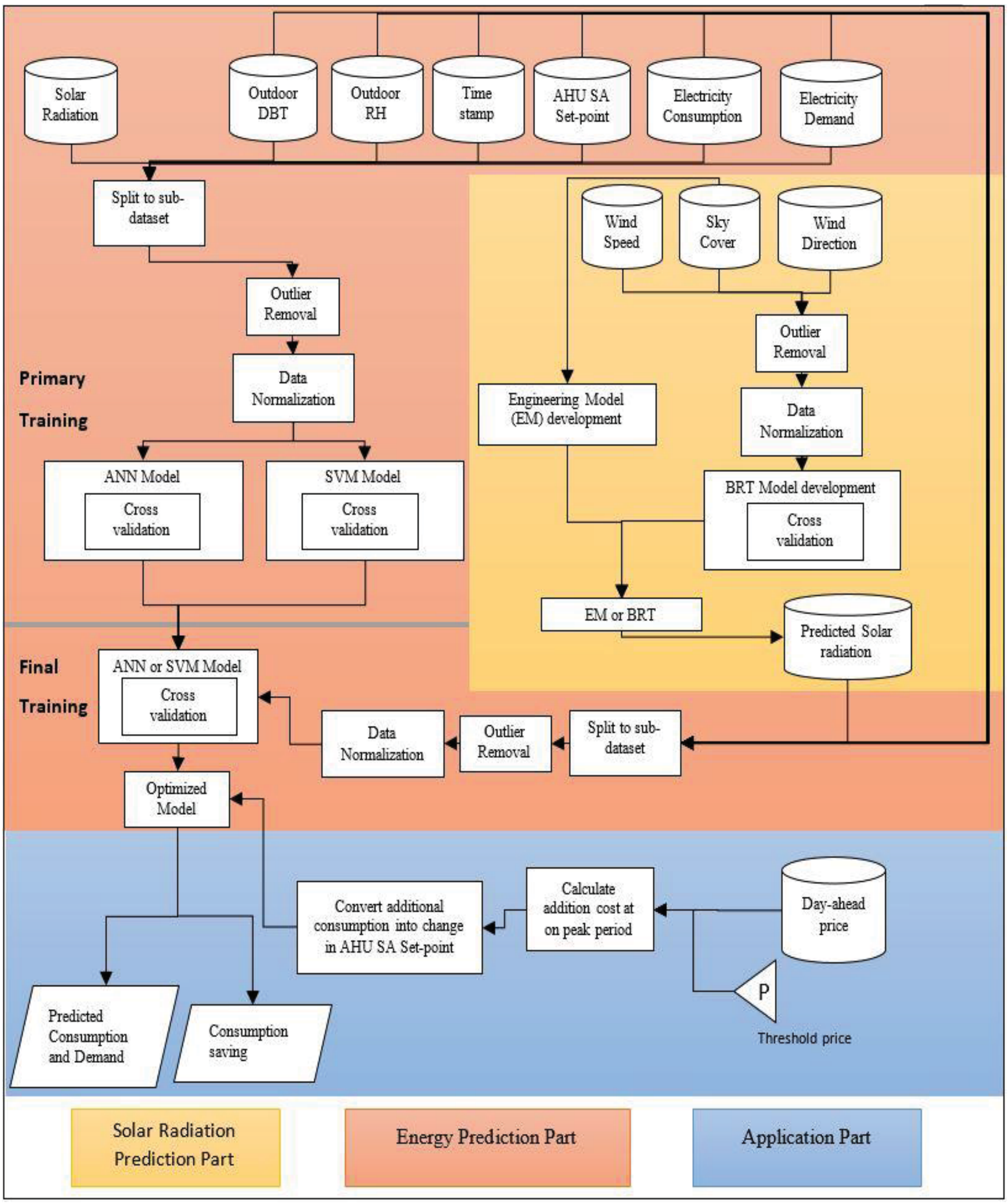

Fig. 1 Schematic outline of the research

Last, the processes in blue are associated with the application part. A proposed DR algorithm is introduced in this part as an application example of demand prediction. The algorithm can control the AHU supply air set-point; those changed set-points will then be input into the prediction model again and the level of output demand will be deduced from this operation. The hypothesis in this part is that the historical DAP is assumed as the price the occupant actually paid.

C. Machine Learning Algorithm

1. ANN

ANN is one of most high-end and accurate machine learning methods to find the inherent complex relationship among input features. As an effective model, ANN is a common tool in the building energy prediction issue. In this study, the 'build network' in Pybrain [16] is used to predict the day-ahead energy consumption and peak demand.

ANN is a kind of multiple perceptron network (MPN) which consists of three layers: input layer, hidden layer and output layer. The input layer only contains the input features of the model; the hidden layer consist of processing nodes which are called perceptrons (referred to in some papers as neurons); output layer consist of perceptron, which numbers are identical with the number of output. Each perceptron in ANN contains the activity function and weight, which means that each perceptron receives input from other the perceptrons and multiply its weight, and then apply it to the activities function, as shown in:

$$a_i^{(L)} = g(\sum_{i=0}^{l} \Theta_{(l-1)i}^{(L)} x_i) \tag{1}$$

where the $a_i^{(L)}$ is the output of perceptron i in layer L, the $\Theta_{(l-1)i}^{(L)}$ is the matrix of weights controlling function mapping from layer $L-1$ to L, and g is the activities function. In common, the activities function is a sigmoid function, which is given here:

$$g(x) = \frac{1}{(1+\exp(-x))} \tag{2}$$

In a certain FFNN model, the activities function for all perceptrons are identical, but matrix of weights varies. How to determine the matrix of weights is the most important part during the model training process. Matrix of weights can directly affect the performance of the model. In Pybrain, the initial weights are randomly determined. Then the gradient descent method is used to optimize the matrix of weights. The basic principle of gradient descent is minimizing the distance between actual outputs and predicted one:

$$\Theta \rightarrow \arg \underbrace{min}_{\Theta} \sum (y_i - f(x_i))^2 \tag{3}$$

where y_i is the actual output, and $f(x_i)$ is the output calculated from the model, then the distance is minimized by the gradient descent method:

$$\nabla E[\vec{\Theta}] = (\frac{\partial E}{\partial \Theta_1}, \frac{\partial E}{\partial \Theta_2}, \frac{\partial E}{\partial \Theta_3}, \frac{\partial E}{\partial \Theta_4} \cdots \frac{\partial E}{\partial \Theta_l}) \tag{4}$$

Then, each weight is optimized by the training rule shown below:

$$\Delta\Theta_i = -\eta \frac{\partial E}{\partial \Theta_i} \tag{5}$$

where η is the learning rate in each iteration. After getting the error $\Delta\Theta_i$, the model would iteratively backtrack this error from output layer to the hidden layers and then optimizing the weights. After a certain number of iterations or results approaching some coverage criteria, iteration will cease, and it could be assumed that the excepted matrix of weights is found.

2. SVM

SVM is the most powerful machine learning method. In last decades, SVM has become a common tool used in building energy prediction.

The first step of SVM is that the input is mapped into the feature space by certain types of kernel, and then the hyperplane is constructed to get the result. In general, the relationship between input and output in SVM is:

$$y = f(x) = w\phi(x) + b \tag{6}$$

where w is the orthogonal vector to the separating the hyperplane, and $\phi(x)$ is the nonlinear high dimension feature space. The margin can be represented by the function:

$$m = \frac{2}{\|w\|} \tag{7}$$

In order to find the maximum margin, the function can be converted to a quadratic programming optimization problem:

$$\text{minimize} \quad J(w) = \frac{1}{2}\|w\|^2 \tag{8}$$

$$\text{constrained to} \quad \begin{cases} y_i - w^t x_i - b \geq \varepsilon \\ w^t x_i + b - y_i \geq \varepsilon \end{cases}$$

However, some data would make the f(x) not exist. If we want to allow this kind of error, slack variable ξ_i and ξ_i^* can be introduced here to deal with this infeasible constraints of the optimization. So, the above function can be rearranged to:

$$\text{minimize} \quad J(w) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n} I(\xi_i + \xi_i^*) \tag{9}$$

$$\text{constrained to} \quad \begin{cases} y_i - w^t x_i - b \geq \varepsilon + \xi_i^* \\ w^t x_i + b - y_i \geq \varepsilon + \xi_i^* \\ \xi_i^*, \xi_i > 0 \end{cases}$$

The constant C represents the trade-off between the shape of the function and the training error larger than ε are tolerated. Then the Lagrange function and kernel function $k(x_i x_j)$ are constructed in both the objective function and corresponding constraints, the equation above can also be expressed as:

$$\text{maximum} \quad \begin{cases} -\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*)k(x_i x_j) \\ -\varepsilon\sum_{i=1}^{n}(\alpha_i - \alpha_i^*) + \sum_{i=1}^{n} y_i(\alpha_i - \alpha_i^*) \end{cases} \tag{10}$$

$$\text{constrained to} \quad \begin{cases} \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \in [0, C] \end{cases}$$

where α_i is the Lagrange multipliers. Then function f(x) can be written as:

$$y = f(x) = w\sum_{i=1}^{n}(\alpha_i - \alpha_i^*)\, k(x_i x_j) + b \tag{11}$$

The final step is to choose a suitable kernel function that can find the strong relationship between input feature and hourly electric consumption. Kernel functions include linear kernel, polynomial kernel, Sigmoid kernel and radial basis function kernel (RBF kernel). The RBF function is selected in this research for its high performance to cope with non-linear data, this function is:

$$k(x_i x_j) = \exp(-\frac{1}{2\sigma^2}\|x_i - x_j\|^2) \quad (12)$$

where σ is the free parameter, it needs to be determined in the work of next stage.

3. Zhang-Huang Solar Model

Zhang-Huang Model is a model to predict the solar radiation by total sky cover, dry bulb temperature, relative humidity and wind speed. Initially, this model is developed only for the China TMY file. However, according the description from the EnergyPlus engineering reference [17], this model is also suitable to other locations as well. The formula of this model is:

$$\mathrm{I} = \frac{[I_0 \times \sin(h) \times (c_0 + c_1 CC + c_2 CC^2 + c_3(T_n - T_{n-3}) + c_4 \times \varphi + c_5 \times V_w) + d]}{k} \quad (13)$$

where I is the estimated hourly solar radiation, I_0 is the global solar constant, h is the solar height angle, CC is the sky cover, φ is relative humidity, T_n, T_{n-3} is outdoor dry bulb temperature at current stage (n) and previous stage (n-3), V_w is the wind speed, $c_0, c_1, c_2, c_3, c_4, c_5, d, k$ are the regression coefficient.

$$c_0 = 0.5589, c_1 = 0.4982,$$

$$c_2 = -0.6762, c_3 = 0.02842,$$

$$c_4 = -0.00317, c_5 = 0.014,$$

$$d = -17.853, k = 0.843$$

The historical solar height angle (h) can be calculated by the following formula:

$$\sin(\mathrm{h}) = \sin(\phi)\sin(\delta) + \cos(\phi)\cos(\delta)\cos(\mathrm{HRA}) \quad (14)$$

where ϕ is local latitude, δ is the current declination of the Sun, HRA is the hour angle in local solar time.

4. Boosting Regression Tree (BRT)

According to the research of Jie Zhao [19], BRT is an effective model to predict solar radiation. It consists of an ensemble of regression trees by boosting method. This part of the algorithm is implemented by the 'Decision Tree Regressor' and 'AdaBoost Regressor' module in Scikit-learn.

Regression tree identifies regions by the most homogeneous responses to the prediction result, to split them into different a rectangular space. Each sample would fit a constant which is the mean response for the observation in each rectangular space. The attribution in each level of a regression tree is determined by the largest standard deviation reduction. Standard deviation reduction is based on the decrease in standard deviation before and after dataset split by an attribution.

$$\pi^* = \mathrm{avgmax}(\mathrm{S(t)} - \mathrm{S}(\mathrm{t}, \pi^i)) \quad (15)$$

$$\mathrm{S}(\mathrm{t}, \pi) = \sum_{i=1} \sum P(t, \pi_i) S(t, \pi_i) \quad (16)$$

where π^* is the most suitable attribution in current level of a regression tree, π^i is the arbitrary attribution in this case, and $\mathrm{S}(\mathrm{t}, \pi)$ is the overall standard deviation from all sub-datasets which is split by attribution π. $P(t, \pi_i)$ is the fraction of one sub-dataset which split by attribution π, and $S(t, \pi_i)$ is the standard deviation of one sub-datasets which split by attribution π. However, single regression tree only can assign samples into corresponding rectangular spaces. The prediction results of samples in one rectangular space are all the same; it may lower the model accuracy.

Boosting is a method to improve the performance of regression tree, based on the idea to find and combine many simple models, rather than to find one complicated model. Although many other techniques, such as bagging, stacking can combine the results from different models; however, boosting is the most efficient because it is forward procedure, not a backtracking one. The aim of BRT is to minimize the loss function for each tree. For example, the second regression tree is fitted to the residuals of the first tree, and the residual of the second tree is the calculated and fitted by the third regression tree and so on. Therefore, the final model is a linear combination of all these regression trees.

D.Application: DR Algorithm

The demand control used in this study is a peak reduced algorithm. Peak reduced system is designated to reduce the electricity demand at the on-peak period by changing the AHU supply air set-point (AHUSP). In general, the control strategy of the proposed DR algorithm can be divided into four steps: (1) this system will have a threshold electricity price P_{th}, which is used to determine whether the current electricity price is at peak or not. This threshold price can be modified according to the occupancy preference; (2) if P_{th} is more expensive than the current price P, the purposed DR algorithm starts to change the AHUSP; (3) the difference between the threshold and current price is used to calculate the additional expense; then this additional expense should be offset by new AHUSP. Finally (4), DRS will re-predict the electricity demand based on the change of AHUSP, this new electricity demand can represent the performance of this DRS algorithm. However, AHUSP in the 2015 historical data basically only has three values (86 ^{o}F, 75 ^{o}F and 65 ^{o}F). Such small diversity is not suitable to train the machine learning model.

To solve this problem, the ideal solution is changing building AHUSP to increase its diversity and recording the data again. It is not practical in the current condition; collecting data is a long term process lasting at least one year. However, we found that the AHU supply air temperature have a high dispersion degree. AHU supply air temperature and AHUSP are shown in Figs. 2 and 3. Therefore, we use the AHU supply air temperature, and assume it as AHUSP. A clarification stated here in order to avoid any confusion, is that AHUSP mentioned in the following sections also means AHU supply air temperature.

Price is the only one variable to determine whether it is on-peak period. Therefore, we assume the user of this algorithm is a participator in the wholesale market. However, it is too complicated to simulate the price by predicted demand, so the historical DAP of one node will be used here. Since the target building in this research is located in Pittsburgh, Pennsylvania, 2015 DAPs for a node for the Duquesne Light (DUQ), which is the local utility, is assumed as the DAP match predicted demand. Two researches conducted by Yoon [6] also made the similar assumption. In his study, the DAP or RTP are assumed as the retail electricity price, respectively, because many utilities use both types of pricing to build their own DR system.

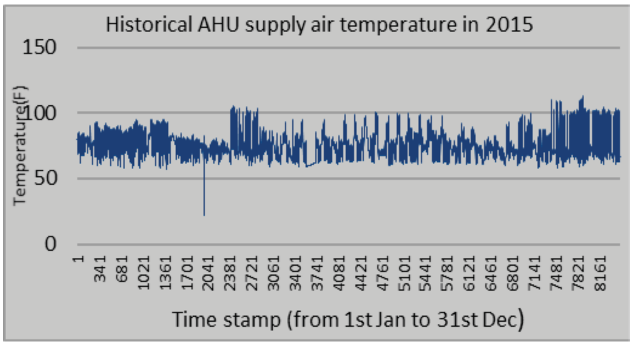

Fig. 2 Historical AHU supply air temperature in 2015

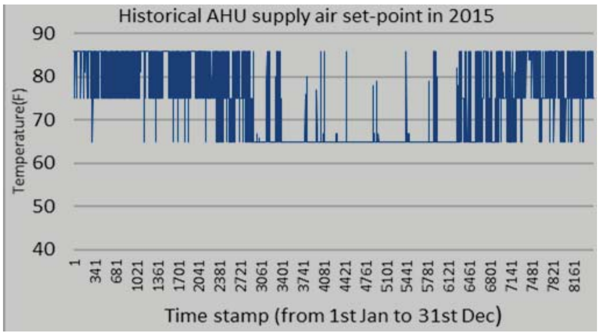

Fig. 3 Historical AHU supply air set-point in 2015

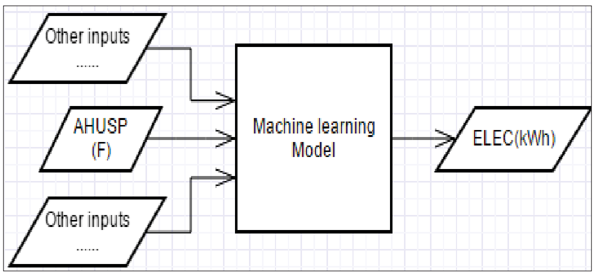

Fig. 4 Electricity demand calculated by proposed machine learning model

Then, the algorithm is described in detail. The equation for electricity demand (ELEC) is the black box model—proposed demand prediction model attained from model selection(f()). The only one dependent variable used here is AHUSP:

$$\mathrm{ELEC[kW]} = \mathrm{f(AHUSP)} \tag{17}$$

The threshold electricity price (P_{th}) is the trigger for this algorithm, the default value of (P_{th}) is 0.035 \$/kWh. It means that, during the off-peak period, the DAP less than 0.035 \$/kWh, while during the on-peak period, the electricity price is beyond 0.035 \$/kWh. Therefore, the difference between the current price (P) and the threshold is:

$$\Delta\mathrm{P[\$/kWh]} = \mathrm{P} - P_{th} = P - 0.035 \tag{18}$$

The next parameter aimed to figure out is the coefficient ($\alpha_{t/\$}$) of temperature by financial saving. It can be interpreted as how much degrees set-point (AHUSP) change can result in one dollar saving.

$$\Delta\mathrm{T[°F]} = \mathrm{k} \times \alpha_{t/\$} \times \mathrm{ELEC} \times \Delta\mathrm{P} \tag{19}$$

This k is the changing gradient, the coefficient (α_t) can be calculated by the threshold electricity price (P_{th}) and the coefficient ($\alpha_{t/e}$) of temperature by electricity saving.

$$\alpha_{t/\$}[°\mathrm{F}/\$] = \frac{\alpha_{t/e}}{P_{th}} \tag{20}$$

The coefficient ($\alpha_{t/e}$) of temperature by electricity saving can be obtained from the electricity prediction model. In order to prevent the thermal discomfort, the maximum range set-point change is limited within ±3°C, which is equivalent to ±5.6 °F. This 3 °C was determined according to ASHRAE 55 thermal comfort range. Therefore, the equation of the $\alpha_{t/e}$ is shown below:

$$\alpha_{t/e}[°\mathrm{F/kWh}] = \frac{5.6+5.6}{f(AHUSP+5.6)-f(AHUSP-5.6)} \tag{21}$$

In order to make this target building suitable for this DR algorithm, three assumptions are proposed here: 1. CSL building owners purchase electricity from the wholesale market directly, rather than using their on-site energy source or purchasing from the retail market; 2. The historical DAP of a local node (Duquesne Light) is assumed as the actual price paid by building owner; 3. AHU supply air temperature appears as the AHU supply air set-point.

III. Development of Prediction Models

A. Data Preparation and Normalization

In this study, all kinds of data are acquired for the year 2105 at one-hour time intervals, for a total of 8,670 data points. After excluding unidentified data and outliers, 8,402 available data points remain.

Because building operation and occupancy conditions vary from time to time, there is no evident relationship between output and input over an annual dataset. In order to show those patterns clearly, the entire year's data should be separated into two sub-datasets by different types of HVAC operation. Fig. 5 shows the plot of output versus input when HVAC heat or cool a building. In addition, the consumed energy profile would also be significantly different during weekdays and the weekend.

Intuitively, at similar weather conditions, consumption will be much higher during the weekdays than that at the weekend. Therefore, in this research, the data for an entire year is separated into 4 sub-datasets according to the heating or cooling season and weekday or weekend. The four training datasets are shown in Fig. 6. Also, data normalization is needed before importing features into model, which is used to standardize the range of features. The Gaussian normalizer, which can more handle the extreme data out of range of training data, is used in this study.

$$x_{new} = \frac{x_{original} - \bar{x}}{\sigma} \tag{22}$$

where $x_{original}$ is original data, $\bar{x}$ is the mean of that feature, σ is the standard deviation of that feature, x_{new} is the standardized data. Then, during the training, leave group out cross validation would be used to prevent the over-fitting issue.

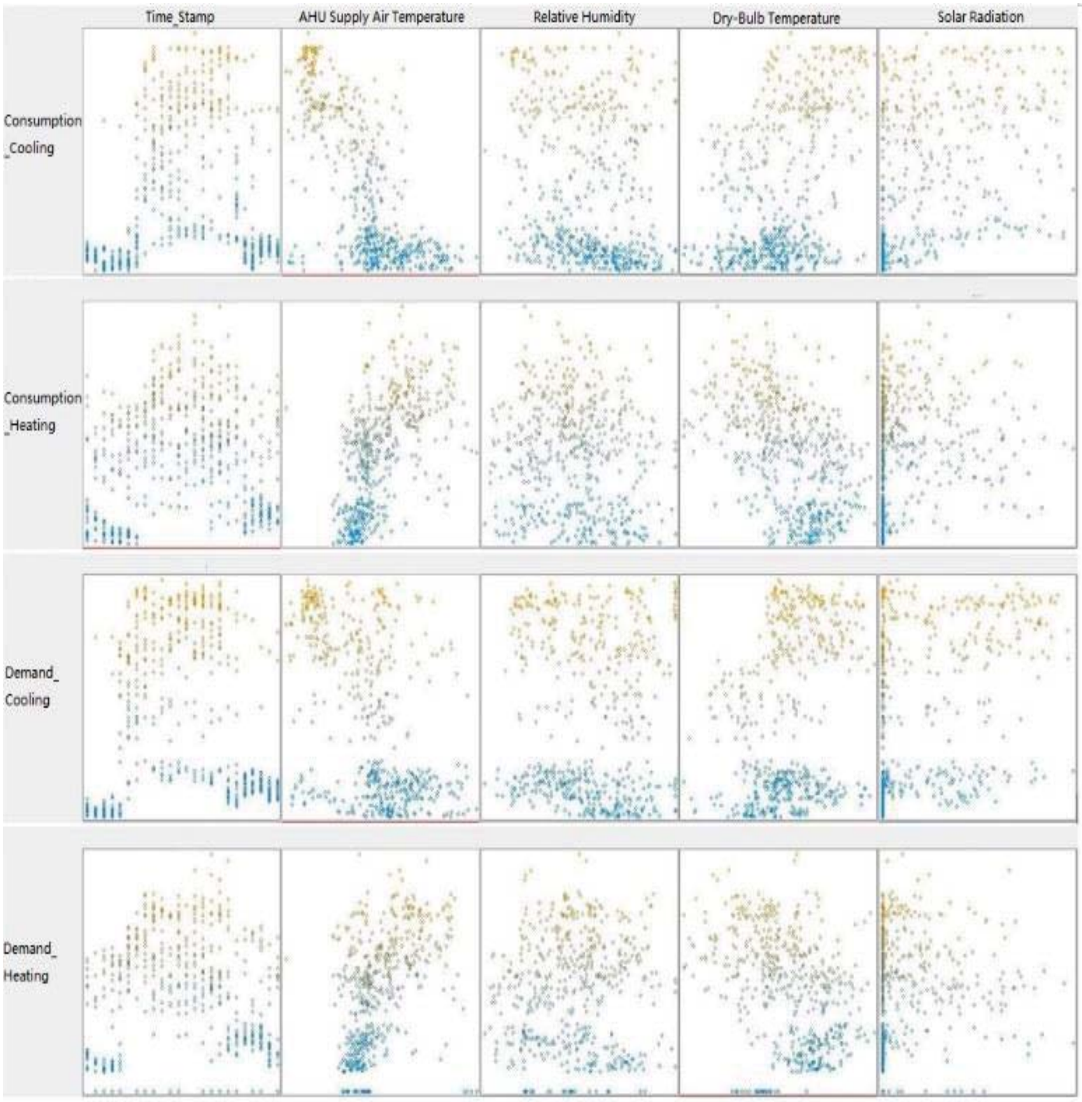

Fig. 5 Input versus Output Plot in heating and cooling

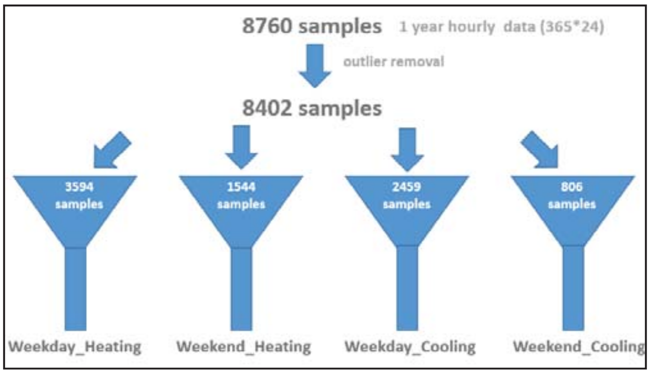

Fig. 6 Manually cluster into sub-dataset

B. Model Selection

1. BRT

Many parameters influence the performance of BRT, and here we only examine two parameters: the number of estimators and the maximum depth of the decision tree. The best combination of numbers of estimators and the maximum depth will be determined by the grid search, which can exhaustively search all parameter combinations in the parameter grid.

The range of numbers of the estimators examined here is [50, 100, 150, 200, 250, 300, 350], the range of maximum depth examined here is [4, 6, 8, 10, 12, 14, 16, 18].

2. SVM

The SVM performance mainly depends on kernel types, penalty term, kernel coefficient (gamma) and width of e-tube (epsilon). Previous studies shows that the first three parameters [8], [20] have a significant impact on the result: penalty term, kernel coefficient (gamma) and grid search also used in the SVM. The range of C in the grid search is from 2^2 to 2^8, the range of grid search is from 2^{-8} to 2^{-1}, the kernel type is RBF.

3. ANN

The training principle of ANN is gradient descent, so the grid search is unavailable for ANN training. The ANN model selection is implemented by the maximum iteration number, which is 350.

C. Performance Criteria

The performance criteria are the index to evaluate the performance of the machine learning model; it can estimate the overall prediction error (or accuracy) between the predicted value and the actual value. The best model must have the least error; different performance criteria can have a different way to define this error. In the energy prediction model, the performance criteria always used are the coefficient of variance of the root mean square error (CVRMSE). To compare with other performance criteria, the CVRMSE can directly represent the overall error between predicted data and real data in a model. The ASHRAE Guideline 14-2002 [21] also recommends CVRMSE as the performance criteria in an energy model. For whole building hourly prediction, the ASHRAE Guideline 14 requires CVMRSE are less that 30%. The equation of CVRMSE is shown below:

$$\text{CVRMSE} = \frac{RMSE}{\bar{Y}} \times 100\% \tag{23}$$

$$\text{RMSE} = [\text{MSE}]^{0.5} = [\frac{\sum_{i=1}^{n}(Y_i-\hat{Y})}{n-p}]^{0.5} \tag{24}$$

where $\hat{Y}$ is the predicted value of output, $\bar{Y}$ is the average of output, Y_i is the real value of output, n is the number of data sample and p is the number of model features. The smaller the value of CVMSE, the closer the regression is predicted to real data.

In the solar radiation prediction model, the performance criteria used is r^2, also referred to as the coefficient of determination. It is an important criterion in regression analysis. It means that the amount of the proportion of the predicted value can be explained by the actual values. The equation of r^2 is:

$$r^2 = \frac{\sum_i(\hat{Y}-\bar{Y})^2}{\sum_i(Y_i-\bar{Y})^2} \tag{25}$$

where $\hat{Y}$ is the predicted value of output, $\bar{Y}$ is the average of output, Y_i is the real value of output. The range of r^2 is [0,1], where 0 means randomly distributed and 1 means prefect prediction.

IV. Results

Due to the use of multiple models, the results are presented in this section to avoid any confusion among each model, and identifiers are assigned to the corresponding model in Table I.

SVM, BRT use grid search for model optimization. In grid search, models constructed by all possible parameter combinations would be trained one by one, so the model with the smallest CVRMSE during grid search is used to represent the training performance for the SVM and BRT model. Whereas, due to ANN trained by iteration (gradient descent), we use the average last 10 iterations of CVRMSE to represent the training performance of ANN, rather than the smallest value.

A. EM1

The training performances of the SVM model and of the ANN model in EM1 are shown in Fig. 7. SVM have not met the requirements as set out in the ASHRAE GUIDLINE 14 [21] that CVRMSE larger is greater than 30% in two datasets (weekend_heating and weekday_cooling). Whereas, all CVRMSE from ANN are around 20% and all are less than 30%. Obviously, ANN has better training performance in electricity consumption prediction, and the ANN method would be used in EM3.

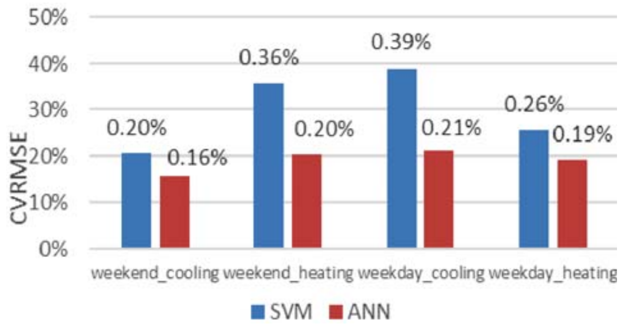

Fig. 7 Comparison of CVRMSE from SVM ANN in EM1

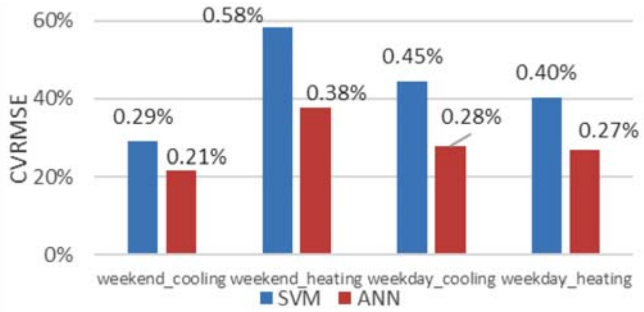

Fig. 8 Comparison of CVRMSE from SVM ANN in EM2

B. EM2

The comparison between the CVRMSE of SVM and that of ANN is shown in Fig. 8. Just like EM1, ANN has a smaller CVRMSE than SVM in all four datasets. Therefore, ANN would be used in EM 4. However, although there have been dramatic improvements by ANN rather than SVM in the weekend_heating dataset (from 58.43% to 37.65%), CVRMSE is still higher than 30%.

C. SRM

SRM aims to find a more suitable model between the Zhang-Huang method and BRT for solar radiation prediction. In the Zhang-Huang method, no optimization process needed. The annual result is shown in Fig. 9. The coefficient of determination for the whole year prediction is 0.522.

In terms of the BRT method, the grid search is also used here to find the best parameter combination. The results of the grid search are shown in Fig. 10. The best training performance occurred at 150 estimators and 18 maximum depth, the coefficient of determination is 0.677.

TABLE I
IDENTIFIER OF DIFFERENT MODELS

Identifier	Method	Input	Comments
EM1(consumption prediction)	ANN and SVM	Time stamp, historical dry bulb temperature, relative humidity, AHU supply air set point, solar radiation.	This model is used to find which method has a better performance for consumption prediction, the better one is chosen in the following part.
EM2 (demand prediction)	ANN and SVM	Time stamp, historical dry bulb temperature, relative humidity, AHU supply air set point, solar radiation.	This model is used to find which method has a better performance for demand prediction, the better one is chosen in the following part.
SRM (solar radiation prediction)	Zhang-Huang and BRT	Historical dry bulb temperature, relative humidity, sky cover, solar height angle for Zhang-Huang method Historical dry bulb temperature, relative humidity, wind speed, wind direction, day, month for BRT method	This model is used to find which method has a better performance for solar radiation prediction, the better one is chosen to predict solar radiation as input in Energy Models 3 and 4
EM3 (consumption prediction)	ANN or SVM (the better method to get from Energy Model 1)	Time stamp, historical dry bulb temperature, relative humidity, AHU supply air set point, and predicted solar radiation generated the better method from the Solar Model.	This model is the final model to predict consumption
EM4 (demand prediction)	ANN or SVM (the better method get from Energy Model 2)	Time stamp, historical dry bulb temperature, relative humidity, AHU supply air set point, and predicted the solar radiation generated by the better method from the Solar Model.	This model is the final model to predict the demand

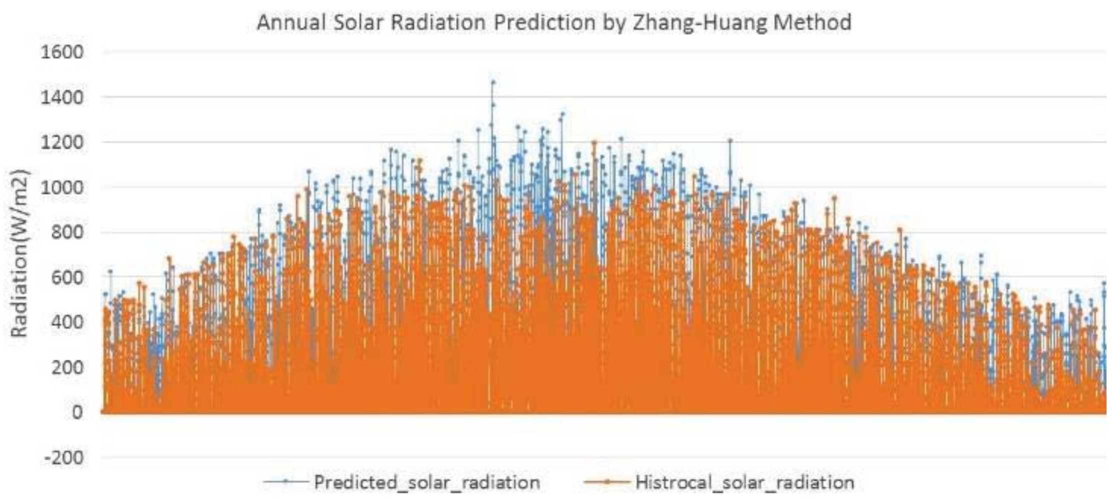

Fig. 9 Annual Solar Radiation Prediction by Zhang-Huang Method

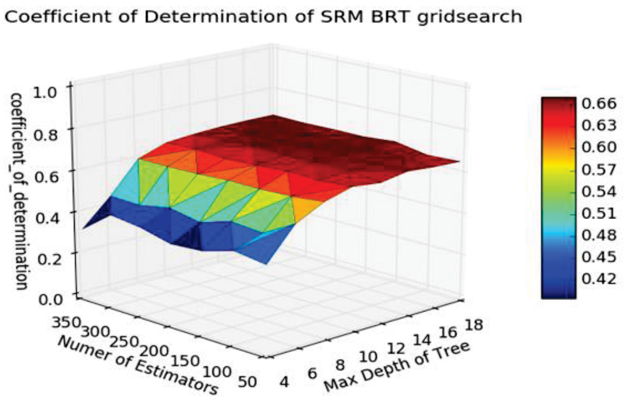

Fig. 10 Coefficient of Determination of SRM BRT Grid Search

D.EM3

According to the results of EM1, ANN is implemented in EM3. This model is used to examine whether the predicted solar radiation attained from SRM can be used as an input. With the exception of solar radiation, other inputs remain the same as the inputs in EM1. The ANN results between EM3 and EM1 are shown in Fig. 11.

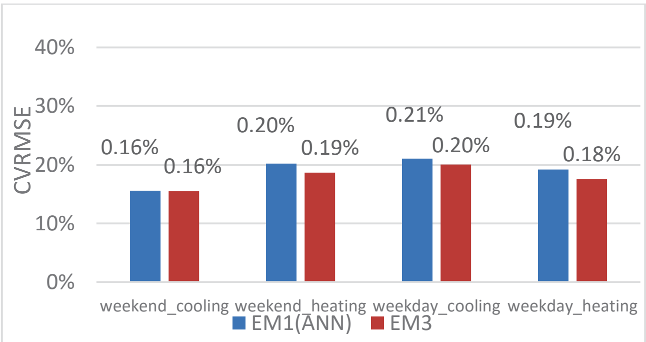

Fig. 11 Comparison of CVRMSE from ANN in EM1 and EM3

After comparison, the 2015 dataset was again used as test dataset to represent the whole year testing performance. The results are shown in Table II. The testing predicted the results even more accurately than the ANN results shown in Fig. 11.

TABLE II
EM3 TESTING PERFORMANCE

CVRMSE	Weekday	Weekend
Cooling	19.87%	13.23%
Heating	17.01%	18.67%

E. EM4

According to the results of EM2, ANN is implemented in EM4. EM4 is used to examine whether solar radiation predicted from SRM can be used as an input. With the exception of solar radiation, other inputs remain the same as the inputs in EM2. The ANN results between EM4 and EM2 are shown in Fig. 12.

After a comparison between EM4 and EM2, the results again used 2015 dataset as the test dataset are shown in Table III. The CVRMSE obtained from weekend_heating in Fig. 12 and Table III are quite different and are 32.42% and 57.31%, respectively. And for weekday_heating from III, the CVRMSE becomes beyond 30%, which represents an increase from 28.49% to 37.49%.

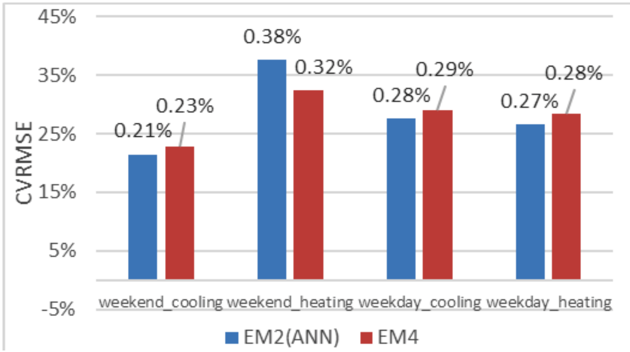

Fig. 12 Comparison of CVRMSE from ANN in EM2 and EM4

TABLE III
EM4 TESTING PERFORMANCE

CVRMSE	Weekday	Weekend
Cooling	27.91%	22.29%
Heating	37.49%	57.31%

F. Total Electricity Saving from DR Algorithm

In this part, the proposed DR algorithm is implemented in EM4 to show how much energy and how much money can be saved. The current price assumed is the DAP acquired from PJM DataMiner [22]. Inputting DAP data and EM4 model into the DR algorithm, the AHU set-point change is shown in Fig. 13, the demand reduction and energy saving results are shown in Fig. 14. However, the results of this DR algorithm also reveal two main problems: Fig. 13 shows the demand increase even at certain time stamps (there are some negative values in the demand reduction diagram); the change in the AHU set-point is not uniform and seems like a random increase or decrease

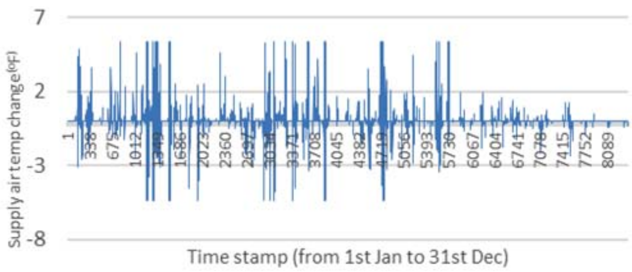

Fig. 13 AHU supply air set-point change by DR algorithm

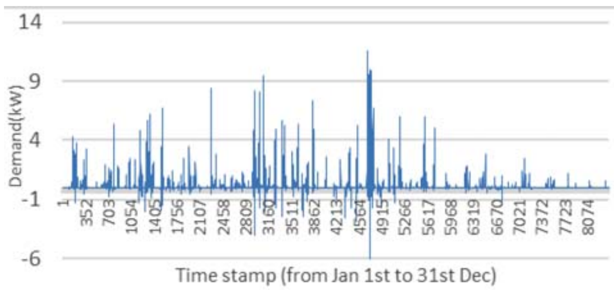

Fig. 14 Demand reduction in each hour by DR algorithm

V. Discussion

A. Model Performance

1. Model Performance

Firstly, the consumption (EM1) and demand prediction (EM2) both show that the ANN has the better training performance than SVM. Although some literatures relevant to the topic indicate that SVM is more frequently used than ANN, each case is identical and it is very difficult to conclude a best method. In most of studies they are both black boxes and unable to be interpreted.

Most of the literature review is based on monthly or daily predictions, rather than hourly prediction. The usage pattern is more difficult to predict when time interval becomes smaller. In a small time interval, the electricity demand may be dominated by a specific usage. However, input in this research, such as dry bulb temperature, and time stamp are quite general. There must be some unknown relationship between these input features. ANN can calculate the hidden relation between the inputs, which is a potential reason for why ANN has the better training performance.

Secondly, comparing EM1, EM2 and EM3, EM4, the results show that CVRMSE of consumption prediction is smaller than that of demand prediction. Therefore, hourly consumption is easier to predict than hourly peak demand. During the building system operation, some loads only occurred in a certain moment, rather operating over a long time. For instance, the load of an elevator is considerable when it is in operation, but the operating of elevator only lasts between 10 and 20 seconds at one time. The total load would drastically increase within a few seconds of the elevator starting operation. Therefore, the operation of this kind of equipment in a building would greatly affect the maximum demand in each hour. However, this type of short-run equipment may have a relatively small influence towards hourly consumption. Table IV also shows that the standard deviation of the hourly peak demand is larger than that of consumption. In other words, the dispersion degree of the hourly peak demand is higher than that of hourly consumption, thus demand is more unpredictable than consumption.

TABLE IV
STANDARD DEVIATION OF HOURLY PEAK DEMAND AND CONSUMPTION

	Hourly peak demand	Hourly consumption
Standard deviation	10.89	9.25

Thirdly, to consider the training performances comparison between the predicted solar radiation trained model and that of the actual solar radiation trained model. The comparison between EM1 and EM3 shows that predicted solar radiation can help ANN perform better in consumption prediction; whereas, there is no evident improvement in hourly peak demand prediction (EM2 and EM4). The potential reasons should be: for consumption prediction, solar radiation may be a redundant factor in consumption prediction, or even disturb prediction. It means that in EM1 or EM3, training the dataset without actual solar radiation maybe produce the better result. In contrast, for hourly peak demand prediction, no clues in EM2 and EM4 make it possible to distinguish whether predicted solar radiation performs better. It may result that both actual and predicted solar radiation are not important factors and that neither of them can improve the model.

Fourthly, in a comparison of the testing performance with the training performance, it can be found in some datasets that the testing performance was even worse than training performance. Especially for weekend_heating, where the CVRMSE of the testing even increased from 32.42% to 57.31%; suggesting model under-fitting maybe a potential reason. In order to validate the possible reasons, the maximum iteration number is increased from 350 to 600, and the calculation testing the model performance of weekend_heating was run again. CVRMSE decreased from 57.31% to 28.54%. Obviously, the testing performance is significantly improved from 350 iterations to 600 iterations.

Lastly, in terms of the DR algorithm results, from Figs. 13 and 14, the AHU set-point changed, thus fluctuated, and even had a negative electricity reduction at some time stamps. Both problems should be blamed on the prediction model itself. It means that the model cannot find the precise relationship between energy usage and AHU operation, or the model cannot be simplified as a linear model. Therefore, this model based DR control is not robust.

2. Study Limitations

In this research, the dataset consisted of 2015 full-year data, which was collected from the PI system of CSL and the Pittsburgh local weather station. The first limitation is that the data range is too narrow to contain extreme weather conditions. For instance, if in this year the weather is generally hot than previous years, the model trained with such a dataset would result in prediction bias. The second is that the data source is acquired from different locations, which could decrease prediction accuracy. All data from the PI system are obtained from CSL on-site sensors, but other data, such as sky cover, wind direction are obtained from a local weather station. Although both data sources are in Pittsburgh, they are still different in terms of micro-climate. The third limitation is that the features should be selected in more appropriate way. In this research, all the features, including time stamp, the dry bulb temperature, relative humidity, AHU set-point and solar radiation, are determined based on previous researches. The final limitation is about target building. From the point of view of sustainability, CSL did a good job; however, a too sustainable building brings a few inconveniences.

VI. Conclusion

This research provides a method to develop machine learning models to predict building hourly electricity consumption and hourly peak day-ahead demand. The proposed model enables large building owners to predict their electricity usage patterns. It can help building owners to gain benefits from the electricity wholesale market directly, it can also be used in the DR controller to reduce electricity usage and ease grid stress. The input data are time stamp, outdoor temperature, relative humidity, solar radiation, AHU supply air set-point. Solar radiation is unique among these inputs, which is unavailable for day-ahead prediction. Therefore, the Zhuang-Huang [17] solar model and BRT are first used to predict it. Then, the ANN and SVM are used to predict building hourly consumption and peak demand. From result, BRT can predict a more accurate solar radiation than Zhang-Huang model, ANN performs better than SVM.

Zhang-Huang [17] solar model is empirical model, whereas BRT is a machine learning model. Both of them calculate solar radiation by temperature, humidity, and sky cover etc., from which we can get a day-ahead prediction. As a result, tne Zhang-Huang model can achieve a coefficient of determination = 0.552. BRT performs better, the best result can achieve a coefficient of determination = 0.677. At mean time, all models are trained by historical data (also including historical solar radiation) by both ANN and SVM. For consumption prediction, the CVRMSE range of the historical data trained model are [15.65%-21.03%] for the ANN and [20.46%-38.95%] for the SVM, respectively. For peak demand prediction, CVRMSE range of those historical data trained model are [21.47%-37.65%] for the ANN and [29.01%-58.43%] for the SVM, respectively. Therefore, the ANN is the one that satisfied ASHRAE guideline 14 requirements more, with CVRMSE less than 30%. After that, predicted solar radiation is used as an input in ANN mode. From the result, predicted solar radiation represents a good performance in both hourly consumption and peak demand prediction. CVRMSE trained by predicted solar radiation for consumption and peak demand prediction are [15.50%-20.03%] and [22.89%-32.42%], respectively. All above CVRMSE are training performance with data normalization and cross validation. Then, the models are tested by the same dataset again, its CVRMSE range for consumption and peak demand prediction are [13.23%-19.87%] and [22.29%-57.31%]. From all the results, the CVRMSE of consumption predicted is much lower than that of peak demand prediction; the ranges are also much narrower. It is evident that hourly consumption is much easier to predict and that performance is more stable than hourly peak demand. However, this study also reached the conclusion that the proposed models are not robust to do building control. The result from the proposed DR algorithm indicates that extra electricity consumed even existed during DR algorithm operation. Therefore, the prediction model is not accurate to calculate the electricity saving by changing the AHU set-point.

For future study, dataset improvement and algorithm optimization are two main recommendations. In machine the learning domain, the quality of a dataset is vital prior to building a good model. The data quality also relates to sufficient quantity such as the dataset size, time window of the data and input feature selection etc., all of which may affect the performance of the model, and thus, all should be taken into consideration. In terms of algorithm optimization, the coverage criteria should be carefully set in the ANN; instead rule based clustering, some advanced clustering methods should be used before training data, such as correlation analysis, information gain analysis etc.; also, the building consumption or demand can be split into different parts, and models should be built separately for each part.

ACKNOWLEDGMENT

This research was supported by the authors' supervisor, Azizan Aziz and Bertrand Lasternas, who are appreciated for their cordial and patient guidance.

The authors' would also like to thank all staff in Intelligent Workplace for their much appreciated help and suggestions.

REFERENCES

[1] "Energy Efficiency." Energy Efficiency. Accessed May 03, 2016. http://www.iea.org/aboutus/faqs/energyefficiency/.
[2] Wattles, Paul. "ERCOT demand response overview & status report." In AMIT-DSWG Workshop "AMI's Next Frontier: Demand Response. 2011.
[3] Borenstein, Severin, Michael Jaske, and Arthur Rosenfeld. "Dynamic pricing, advanced metering, and demand response in electricity markets." (2002).
[4] Cochran, Jaquelin, Mackay Miller, Michael Milligan, Erik Ela, Douglas Arent, Aaron Bloom, Matthew Futch et al. "Market evolution: wholesale electricity market design for 21st century power systems." Contract (2013).
[5] Patel, Siddharth, Raffi Sevlian, Baosen Zhang, and Ram Rajagopal. "Pricing Residential Electricity Based on Individual Consumption Behaviors." arXiv preprint arXiv:1312.1243 (2013).
[6] Yoon, Ji Hoon, Ross Bladick, and Atila Novoselac. "Demand response for residential buildings based on dynamic price of electricity." Energy and Buildings 80 (2014): 531-541.
[7] Zhao, Hai-xiang, and Frédéric Magoulès. "A review on the prediction of building energy consumption." Renewable and Sustainable Energy Reviews 16, no. 6 (2012): 3586-3592.
[8] Dong, Bing, Cheng Cao, and Siew Eang Lee. "Applying support vector machines to predict building energy consumption in tropical region." Energy and Buildings 37, no. 5 (2005): 545-553.
[9] Yokoyama, Ryohei, Tetsuya Wakui, and Ryoichi Satake. "Prediction of energy demands using neural network with model identification by global optimization." Energy Conversion and Management 50, no. 2 (2009): 319-327.
[10] QDR, Q. "Benefits of demand response in electricity markets and recommendations for achieving them." US Dept. Energy, Washington, DC, USA, Tech. Rep (2006).
[11] Barbose, Galen, Charles Goldman, and Bernie Neenan. "A survey of utility experience with real time pricing." Lawrence Berkeley National Laboratory (2004).
[12] Federal Energy Regulatory Commission. "Assessment of demand response and advanced metering." (2012).
[13] "Training Material." PJM - Accessed May 03, 2016. http://pjm.com/training/training-material.aspx.
[14] "1.4. Support Vector Machines." 1.4. Support Vector Machines — Scikit-learn 0.17.1 Documentation. Accessed May 03, 2016. http://scikit-learn.org/stable/modules/svm.html#svr.
[15] "Decision Tree Regression with AdaBoost." Decision Tree Regression with AdaBoost — Scikit-learn 0.17.1 Documentation. Accessed May 03, 2016. http://scikit-learn.org/stable/auto_examples/ensemble/plot_adaboost_regression.html.
[16] "PyBrain Documentation." PyBrain. Accessed May 03, 2016. http://pybrain.org/.
[17] Documentation, EnergyPlus. "Engineering reference." (2005).
[18] "V2 Forecast API." The Dark Sky Forecast API Docs. Accessed May 03, 2016. https://developer.forecast.io/docs/v2.
[19] Zhao, Jie. "Design-Build-Operate Energy Information Modeling for Occupant-Oriented Predictive Building Control." (2015).
[20] Li, Qiong, Peng Ren, and Qinglin Meng. "Prediction model of annual energy consumption of residential buildings." In Advances in Energy Engineering (ICAEE), 2010 International Conference on, pp. 223-226. IEEE, 2010.

[21] Guideline, A. S. H. R. A. E. "Guideline 14-2002, Measurement of Energy and Demand Savings." American Society of Heating, Ventilating, and Air Conditioning Engineers, Atlanta, Georgia (2002).
[22] "Data Miner - Locational Marginal Pricing." Data Miner - Locational Marginal Pricing. Accessed May 03, 2016. https://dataminer.pjm.com/dataminerui/pages/public/lmp.jsf.