SmartPeak: Peak Shaving and Ambient Analysis For Energy Efficiency in Electrical Smart Grid

Sourajit Behera
Department of Computer Science
Indian Institute of Technology Patna, India
sourajit.pcs17@iitp.ac.in

Rajiv Misra
Department of Computer Science
Indian Institute of Technology Patna, India
rajivm@iitp.ac.in

ABSTRACT

In modern times, buildings are heavily contributing to the overall energy consumption of the countries and in some countries they account up to 45% of their total energy consumption. Hence a detailed understanding of the dynamics of energy consumption of buildings and mining the typical daily electricity consumption profiles of households in buildings can open up new avenues for smart energy consumption profiling. This can open up newer business opportunities for all stakeholders in energy supply chain thereby supporting the energy management strategies in a smart grid environment and provide opportunities for improvement in building infrastructure with fault detection and diagnostics. In this context, we propose a approach to predict and re-engineer the hourly energy demand in a residential building. A data-driven system is proposed using machine learning techniques like Multi Linear Regression and Support Vector Machine to predict electricity demand in a smart building along with a real-time strategy to enable the users to save energy by recommending optimal scheduling of the appliances at times of peak load demand, given the consumer's constraints.

CCS CONCEPTS

• Information systems → Data analytics • Computing methodologies → Classification and regression trees

KEYWORDS

Energy Management; Smart Buildings; Machine Learning; Household appliances Scheduling; Forecasting

1 INTRODUCTION

Exponential growth of world population complemented with the socio-economic development of the modern day societies technologically has given rise to the usage of electricity as the most widest used form of energy. In recent times, the integration of renewable energy sources and storage systems in the electricity grid has reduced the dominance of non-renewable energy sources towards generation of electricity. But it has in turn introduced certain uncertainties pertaining to exact amount of production of energy from renewable sources. Accurate knowledge about the electricity consumption can lead to achievement of substantial savings in the operating and maintenance costs, increase of reliability of power supply and delivery system and correct decisions for future development. So, the need for switching to a smarter generation and distribution electricity network called smart grid is hailed important to smart sustenance. In some countries buildings contribute heavily to the overall energy demand and can account up to 45% of the total energy consumption of these countries making them the single largest contributor to the total energy consumption [1]. Consequently the research on building energy consumption has been actively pursued by the scientific and industrial community in recent years.With the successful deployment of smart meters across many countries has helped to pile up large sets of valuable information about the building energy consumption. Mining such data can benefit consumers, retailers, utility providers, distributors through important strategies like demand response management, accurate electricity price forecasting, peak demand shifting, energy conservation etc. So the application of sustainable measures and energy efficient models to existing and new buildings will become one of the top priorities for building managers, owners and designers [2][3] in near future. Data-driven systems for building energy consumption modeling approach is effective in discovering hidden algebraic relationships in the dataset abstractly. There has been a paradigm shift from mathematical equation driven system modeling to data-driven insight expansion driven system modeling in many engineering fields [4] specially with the recent boom of the concept of big data [5]. Additionally the application of machine learning algorithms in building energy analysis and forecasting in recent literature [6] [7] has propelled active research in the area of building energy data mining. The application of data mining and machine learning techniques to assess the highly complex consumption data makes it a powerful mechanism for solving the building energy consumption analysis and forecasting problem [8]. In recent times, with the widespread use of smart appliances and integration of the information and communication technology, residential buildings are progressively traversing to be smarter than ever before. This introduces the need to schedule the operation of the smart appliances to maintain the power demand below the peak demand of the building while balancing the consumer satisfaction and comfort. In this context, our aim in the paper is to develop a data-driven system to model and forecast the overall electricity consumption demand of the building using machine learning techniques. The goal of this model is to find the best trade-off between minimization of the smart building energy consumption and appliance scheduling to maintain the consumer comfort and satisfaction. The main contributions in the paper are:

AICCC '18, December 21–23, 2018, Tokyo, Japan

ACM ISBN 978-1-4503-6623-6/18/12.
https://doi.org/10.1145/3299819.3299833

- Propose a data-driven model to forecast the building energy consumption powered by machine learning techniques
- Development of a scheduling algorithm for smart appliances to maintain the electricity demand below the peak building energy consumption while considering the following:
  - consumer satisfaction and comfort
  - operational dynamics of the appliances
  - associated priority of the appliances rated as per consumer comfort

The remainder of this paper is organized as follows: related section is introduced in Section II, a brief introduction of the machine learning techniques deployed is highlighted in Section III. The data, experiment and results of the case-study are discussed in Section IV. The scheduling algorithm and its complexity analysis is explained in Section V before concluding the paper in Section VI.

2 RELATED WORK

In literature, many solutions with focus on machine learning models development and evaluation have been proposed to deal with energy forecasting in smart buildings. Multiple efforts have been made to address reduction of the cost of energy by scheduling of domestic appliances. Several works in literature focus on simultaneous collection and processing of occupancy data since the various operations of a building are related directly or indirectly to the occupant's behavior[9][10]. Recent works have focused on detecting occupants presence through microphones or smart-phones [11]. An Artificial Neural Network(ANN) based prediction of building's optimal scheduling of HVAC load using day ahead prices of electricity and heating usage is presented in [12]. A consumer based behavior and appliance model has been introduced in [13] connecting consumption behavior, socio-economic background data and appliance ownership statistics grouped from surveys for estimation of probability functions which can generate load profiles for a given day. Participants in the great energy shoot out contest used ANN to forecast [14] the total energy demand of a referred building while Decision tree regression was applied in [15] to forecast the electricity usage in different buildings across Hong Kong. Support vector regression was proposed in [16] for predicting monthly electricity bills for four commercial buildings across Singapore. An approach for prediction of load at system level using Support Vector Machine (SVM) has been presented in [17]. A SVM model per hour was trained for each appliance to estimate the electricity demand per hour. Weather inputs and hourly electricity demand from previous two days were given as exogenous inputs to each model. A comparison of Multi-linear regression(MLR), Support vector regression(SVR) and ANN was performed in [18] for electricity demand prediction using simulation data. It was observed that MLR and SVR reported similar prediction accuracy while ANN performed worst amongst the three methods. An approach for predicting the natural gas demand in public buildings using K-nearest neighbor based approach was proposed in [19]. A probabilistic predictive model based upon Bayesian network (BN) for discovering uncertainties among input attributes and quantification of their effects upon system output was proposed in [20]. The approach was tested on the data provided by Pacific Northwest National Lab (PNNL). An approach based upon real-time pricing (RTP) and time of use (TOU) for the autonomous scheduling of appliances in a single home scenario was taken up in [21]. Considering the usage patterns of customers, a scheduling technique based upon dynamic pricing was proposed in [22]. Ample attempts have been made in literature to perform intelligent automatic scheduling of smart appliances using prediction of optimal overall energy consumptions. A scenario comprising of a single-user with knowledge of amounts of energy consumption of all individual appliances in the experimental setting was studied in [23]. In contrast, a scheduling technique on assumption that prior knowledge of operating times of the appliances was presented in [24]. An approach based on neural networks for attainment of optimal scheduling of smart appliances is presented in [25]. A greedy method was used for determining the degree of incorporation of different energy sources in the proposed approach. A smart home based scheduling and load balancing approach for minimization of the energy wastage and achievement of overall improvement of the performance level is proposed in [26]. The operating method of each appliance was modeled using a state machine. The approach was based upon the assumption that domestic electrical appliances behave as random variable with random operations over any time horizon. A load balancing component was introduced for controlling demand of various electronic devices along with load of entire house. The scheduling process operated in addition to various load balancing restrictions while sticking to least slack time (LST) algorithm. Concerning the related work, we have designed and simulated a data-driven system for a smart building to offer energy efficient solution using various collected and managed sensors data.

3 PERFORMANCE ESTIMATION APPROACHES

This section provides an insight into the machine learning techniques which form the gist of the paper, looking in detail the general algorithm namely Multi Linear regression and Support Vector Machine.

3.1 Multi-Linear Regression (MLR)

In a MLR model, a dependent variable (z_i) is defined in terms of linear combination of two or more independent variables (y_{ik}) as represented in Eq. 1

$$z_i = \alpha_0 + \alpha_1 y_{i1} + \alpha_2 y_{i2} + \cdots + \alpha_k y_{ik} + \epsilon_i \tag{1}$$

The coefficients ($\alpha_0, \alpha_1, \cdots, \alpha_k$) in Eq. 1 are predicted using ordinary Least Squares (OLS) for minimizing the residual sum of squares (RSS) represented in Eq. 2

$$RSS = \sum_{i=1}^{n} \hat{\epsilon}_i^2 + \sum_{i=1}^{n} (z_i - \hat{z}_i)^2 \tag{2}$$

where $\hat{z}_i$ signifies the predicted value of dependent variable and $\hat{\epsilon}_i$ represents residual form of the model. It should be kept in mind that the occurrence of specification error in estimation of coefficients of the MLR model can happen because of missing descriptive variables, irregular functional forms and involvement of measurement errors in calculating descriptive variables which can cause a serious bottleneck. For instance, if a model is wrongly-formulated then coefficients of retained variables tend to be inconsistent and biased. In order to be classified as best linear unbiased estimator, MLR model has to fulfill the following requirements:

- each residual satisfies homoscedasticity i.e. same variance with finite value
- the distribution of residuals adhere to a normal distribution with expected value as zero
- no correlation exists between the residuals
- no exact multi-collinearity exists between the explanatory attributes.

If the assumptions regarding homoscedasticity and auto-correlation are violated the predicted coefficients however still remain consistent and unbiased but predicted variances of OLS estimator becomes biased. This ends up causing unreliable computation of t , F test with coefficient of determination of R^2. This advocates the development of the MLR model on principle of *parsimony* and good to fit widely known as adjusted R^2 criteria while maintaining consistency with related theory [27].

3.2 Support Vector Machine (SVM)

SVM, developed by Vapnik [28][29] derives its background from statistical learning theory. Due to its unique structure, the training of SVM model is treated equivalent to solving up of linearly restricted quadratic programming problem which paves the way for a global optimal solution. SVM deals with n data records i.e., $(y_i, Z_i)_{i=1}^n$, with input $y_i \epsilon R^N$ and target $Z_i \epsilon R$. In recent times, SVM is widely applied for solving regression problems to capture the relationship between non-linear inputs to real-valued continuous targets. SVM model utilized for solving regression problems is termed Support Vector Regression (SVR) which has been one of the important data-driven mechanisms for estimating energy consumption of buildings[30]. SVR construction involves creation of a decision function, $F(y_i)$ through a trained using historical data. One important requirement is that for the given input y_i, the output predicted by the decision function must not drift from actual target value Z_i beyond the predefined threshold κ. The decision function in SVR is specified in the form of Eq. 3

$$F(y_i) =< w, \phi(y_i) > +c \tag{3}$$

where $c \epsilon R$ represents the bias . $< \cdot, \cdot >$ and w presents dot-product and weights established within R^N. $\phi(y_i)$ represents a non-linear alignment from input space to a high-dimensional feature space. c and w, the unknown parameters in Eq. 3 which need to be estimated using minimization of regularized risk function [16]. The latter has been solved by introducing a Lagranian L [30]

$$\begin{aligned} L = \frac{1}{2} \parallel w \parallel^2 +d \sum_{i=1}^{n}(\xi_i + \xi_i^*) - \sum_{i=1}^{n}(\eta_i \xi_i + \eta_i^* \xi_i^*) \\ - \sum_{i=1}^{n} a_i(\epsilon + \xi_i - z_i - < w, \phi(y_i) > -c) \\ - \sum_{i=1}^{n} a_i^*(\epsilon + \xi_i^* - z_i - < w, \phi(y_i) > -c) \end{aligned} \tag{4}$$

where $(a_i, a_i^*, \eta_i, \eta_i^*) \geq 0$ represent Lagrange multipliers. $\parallel w \parallel$ represents euclidean norm. $(\xi_i, \xi_i^*) \geq 0$ are the slack variables for copying with infeasible optimization restrictions. The constant $c > 0$ is used for defining trade-off between training error (i.e. over-fitting) and model flatness (i.e. under-fitting). The Lagrange multipliers are independent of each other represented as $\eta_i = d - a_i, \eta_i^* = d - a_i^*$ and value of (a_i, a_i^*) is obtained by solving the dual optimization represented in Eq. 5

$$\begin{aligned} Maximize\ W(a_i, a_i^*) = -\frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} (a_i - a_i^*)(a_j - a_j^*)(\phi(y_i) \cdot \phi(y_j)) \\ + \sum_{j=1}^{n}(a_i - a_i^*) z_i - \epsilon \sum_{j=1}^{n}(a_i + a_i^*) \end{aligned} \tag{5}$$

subject to $\sum_{j=1}^{n}(a_i - a_i^*) = 0$ and $(a_i, a_i^*) \epsilon [0, d]$. With the computed values of a_i, a_i^* the weight w can be re-written as function in terms of $\{a_i, a_i^*, y_i\}_{i=1}^n$. This leads to the formulation of decision function in SVR as represented in Eq. 6

$$F(y) = \sum_{y_i \epsilon SV} (a_i - a_i^*) K(y, y_i) + c \tag{6}$$

where $K(y, y_i) = \phi(y) \cdot \phi(y_i)$. In case of SVR it is labeled as kernel function(s) having various formulas in different applications. Eq. 6 does not consider every input but is restricted to only those support vectors $(y_i \epsilon SV)$ corresponding to
$(a_i - a_i^*) \neq 0$. Additionally the bias term c in Eq. 6 is estimated using support vectors represented in Eq. 7

$$\begin{aligned} c = \frac{1}{N_1}\{ \sum_{a_i\ \epsilon\ (0,d)} [Z_i - \sum_{y_j\ \epsilon\ SV} (a_j - a_j^*) K(y_i, y_j) - \epsilon] + \\ \sum_{a_j^*\ \epsilon\ (0,d)} [Z_i - \sum_{y_j\ \epsilon\ SV} (a_j - a_j^*) K(y_i, y_j) + \epsilon]\} \end{aligned} \tag{7}$$

where N_1 represents the count of support vectors with either
$\{a_i \epsilon (0, d), a_i^* = 0\}$ or $\{a_i = 0, a_i^* \epsilon (0, d)\}$. The SVR model estimates the values for a new input y once, the decision function represented in Eq. 6 is fully specified by the training data.

3.3 Evaluation

Load forecasting involves prediction of future loads. The variation between predicted future load and actual electricity load determines the prediction error. The root mean squared error (RMSE) during any given response time is considered as the evaluation criteria for the prediction accuracy in this paper and represented as Eq. 8:

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T}(P_t - A_t)^2}{T}} \tag{8}$$

where A_t represents the actual load at time t, P_t denotes the predicted load at time t and T denotes the number of days considered in the training or testing dataset.

4 CASE STUDY

4.1 Data Collection

Electricity consumption data from iAWE dataset [31] read by two meters (primary, secondary) from a residential building in Delhi during 2013 is used. Additional sensor data installed in the smart building were captured to play the role of explanatory variables. Electricity consumption was measured at three levels:electricity meter (using Schneider electric EM6400 sensor), circuit panel and appliance level. Power outages are also considered during the recording of the dataset. Uninterrupted supply of water was not available round the clock. To meet the daily needs most of the residents in the building had additional installation of overhead storage tanks. The measurement of the volume of water flow from the supply and usage from the overhead tank has also been considered in the scope of the paper. Voltage was observed to fluctuate from 180V to 260V while the rated voltage was 230V. Additionally, the ambient light and the outside temperature at 1 hour intervals are retrieved from local meteorological office which are assimilated into a final dataset.

4.2 Data Preprocessing

Erroneous data values due to faulty data collection mechanism or misread sensor readings had resulted in negative timestamps. These faulty data records are identified and amputated from the dataset. Power consumption and voltmeter readings close to zero has been rounded off to zero for easier calculations. The 73-day daily per-second scale data sets of intensive and extensive variables are experimented upon for performance evaluation of the proposed system by averaging or summation of the minutely scaled raw input data set. The per-second power consumption data are collected from the smart meters and segmented as follows: $[x_1^1, x_2^1, x_3^1, \cdots, x_{86400}^1]$, $[x_1^2, x_2^2, x_3^2, \cdots, x_{86400}^2], \cdots$ where x_i^k represents the electricity consumption data of the user at i^{th} second on k^{th} day. 86400 represents the number of seconds in a day. The data is then averaged for the calculating the electricity consumption on a per-day basis generating a 73-day average electricity consumption. Furthermore the dataset is averaged for electricity load consumption on the day-of-week scale. The extensive variables like electricity consumption by the residents is summed up and integrated with the intensive variables like outside temperature and ambient light which are averaged and finally combined into a collective dataset.

4.3 Exploratory Data Analysis

The time series variation, frequency distribution, and summary statistics of different variables are analyzed for performance modeling and interpretation of results. Performing summary statistics on the cleansed dataset revealed a variety of intuitive insights comprising peak load demand, average peak load demand, time-interval(s) for peak load demand requests, above average demand or below average demand etc. Peeking at a per-second cross-section of the cleansed dataset recorded on Monday, May 26, 2013 11:17:29 AM on wards represented in Table 1, the peak load was found to be 2271 Watts while the average peak load demand is found to be 1390.2 Watts. Identification of the time-interval for peak load demand for example from timestamp(s) 1301083249 to 1301083291 opens up avenues for application of peak-shaving and power smoothing techniques depicted in the proposed peak-shaving algorithm detailed in section 5. The peak electrical load consumption for the smart building averaged across a week is projected in Figure 1 while the clustered time duration's for the peak demand of electricity across a week is presented in Figure 2. It is observed that the peak consumption averaged for the building lies in the range of [1100 - 1600] watts. Inferring Figure 2, the peak demand on the weekdays occurs in the afternoon mostly after half-day suggesting there is a large usage of devices such as air conditioners during the afternoon because of the high temperature outside leading to high power usage and hence peak consumption in the afternoon. Peak load surge on weekends on time past evening suggests residents of the building engaged in relaxation or recreational activities. The highest power consumption occurs on Wednesday while the least occurs on Thursdays. The above average power consumption averaged out for a week is represented in Figure 3 while the clustered time duration's for the above average power demand of electricity across a week is presented in Figure 4. The above average consumption occurs on Sunday with the large time-interval represented in Figure 4 for weekends suggest the involvement of residents in relaxation purposes. The below average electricity consumption time-intervals

Table 1: Cleaned Augmented Dataset

Timestamp	Power(in Watts)	Temp.(in °C)	Light Intensity(in lux)
1301083249	2019	23	26
1301083291	2271	24	28
1301083320	1431	22	20
1301083372	693	25	32
1301083433	537	21	31

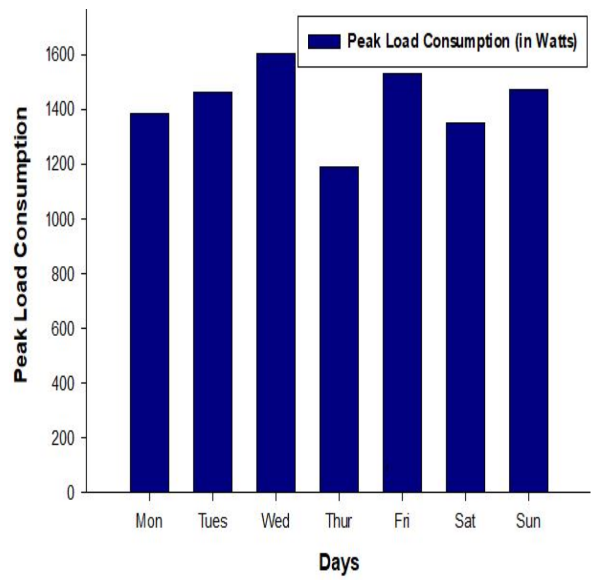

Figure 1: Peak Demand and their respective duration each day

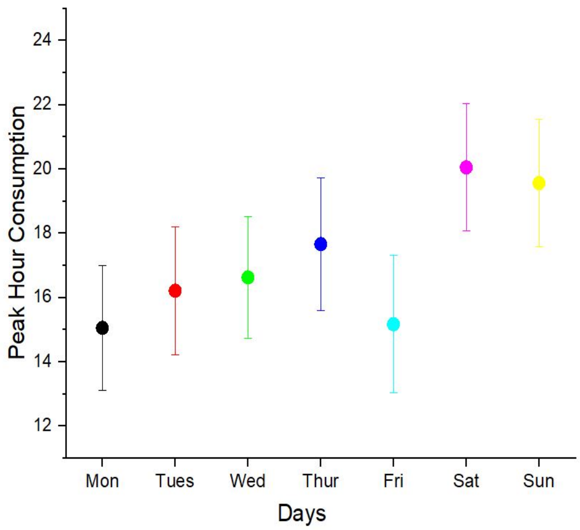

Figure 2: Time intervals for Peak power consumption

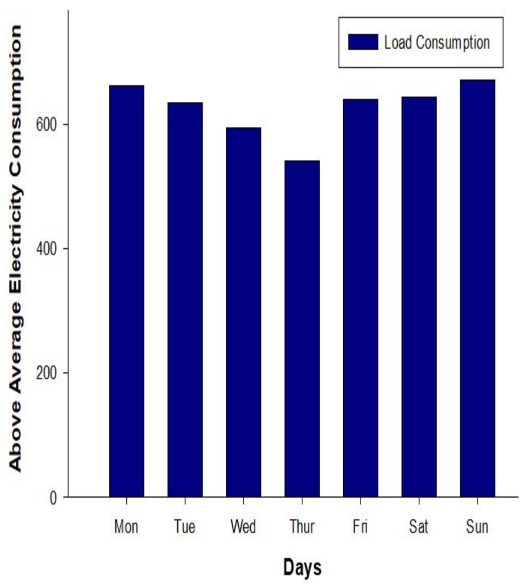

Figure 3: Above average power consumption

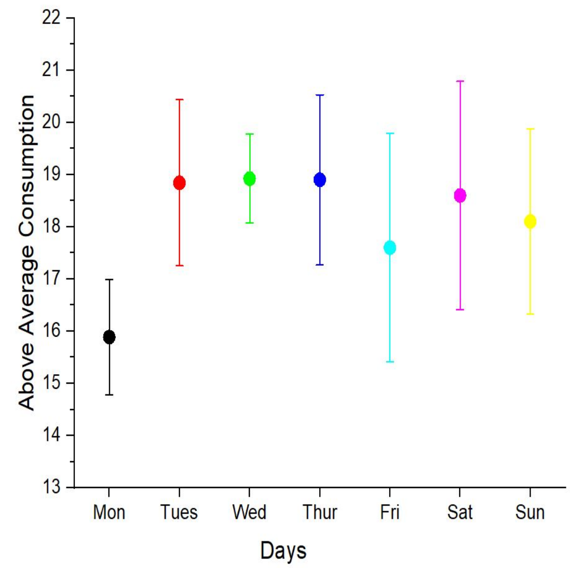

Figure 4: Time intervals for Above average power consumption

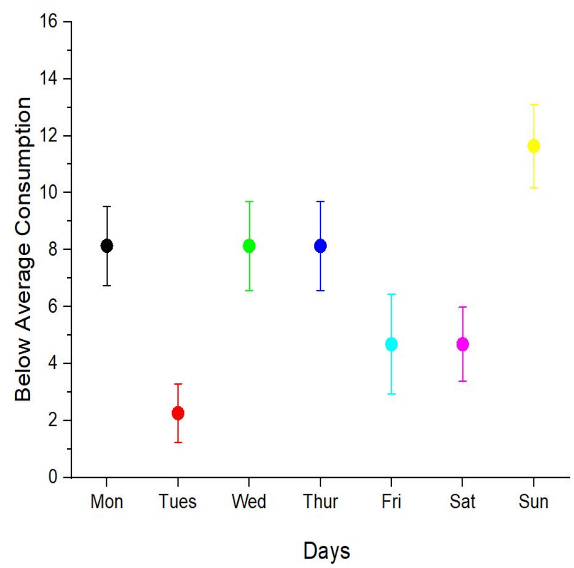

Figure 5: Time intervals for Below average power consumption

presented in Figure 5, suggests that less power usage occurs during the morning as lightweight devices are used. The trend is same for weekdays while differing on Sundays suggesting late starting of day-to-day activities.

4.4 Model Training and Testing

Applying different types of algorithms is intended to build up different models for making comparison and further selecting the best model for this research. Two regression algorithms namely MLR and SVR will be applied for this regression problem of electricity demand forecasting. The total per-day scale cleaned augmented dataset is split into training and testing dataset in a 70-30 ratio. The models are trained using python and scikit-learn library. The range of hyper-parameters of SVR i.e. epsilon and cost are calculated using grid method. Both the selected models of SVR and MLR will are applied on test dataset. The prediction results will be visualized for each one test day of a week, which will indicate the accuracy of the model prediction.

Table 2: Performance Prediction Accuracy of MLR Vs SVR

Day	RMSE(MLR)	RMSE(SVR)
Monday	32.5686 %	12.2396 %
Tuesday	25.9689 %	15.2386 %
Wednesday	29.7897 %	17.2385 %
Thursday	30.7897 %	14.4724 %
Friday	23.2626 %	13.0904 %
Saturday	33.2562 %	12.9796 %
Sunday	19.5965 %	11.5182 %

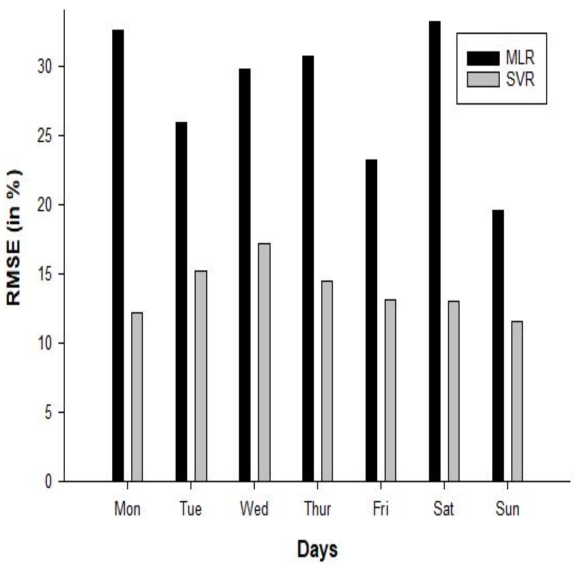

Figure 6: Comparison of Prediction Accuracy of MLR Vs SVR

4.5 Results and Comparisons

The performance of the methods used in this research on testing dataset is presented in Table 2. Fig. 6 represents the a comparison of prediction performance on the smart building dataset using MLR and SVR . Analyzing the results from Table 2 and referring Figure 6, SVR significantly out performs multiple linear regression as it has lower root mean square error for predicting future power consumption. It is seen that the prediction accuracy for both the regression model(s) is best for Sunday while the highest prediction error for load consumption forecasting using MLR is exhibited on Mondays but for SVR it happens on Wednesdays. Even though both of the regression methods are good candidates for electricity loads prediction problem with regular behavior, however SVR is suited better in situations where electricity usage behavior is highly stochastic. Therefore it can be concluded that determining the accurate usage behavior plays an important role for applying the right prediction method.

5 SMARTPEAK

The proposed control scheme for scheduling of home appliances during the duration of a day, the main components along with an overview of the decisions by the controller algorithm are presented in the following sections.

5.1 Appliance Classifications

Home appliances can be divided into two categories as per the appliance type and the customer preferences.

5.1.1 Schedulable Devices: It includes all devices whose power consumption or load profiles can be managed whilst in operational mode. It can be achieved either by reducing their power usage or performing interruptions while still being able to achieve the required energy quota for a day. For instance, the charging rate of an electric vehicle can be controlled in a flexible manner. Given that i^{th}schedulable device denoted by D_i requires w units of electricity represented as $(D_i)_w$, the overall demand of all the n schedulable appliances(Ds) is represented as:

$$Ds = \sum_{i=1}^{n} (D_i)_w \tag{9}$$

5.1.2 Non-Schedulable Devices: It includes all devices who do not exhibit a pre-defined load profile during its operational period and cannot be halted midst operation. For example, washing machines have to stick to their working-cycle of 120 minutes in the rinse-mode without interruption. Given that i^{th}schedulable device denoted by $D_i^{'}$ requires w units of electricity represented as $(D_i^{'})_w$, the overall demand of all the n schedulable appliances(Dns) is represented as:

$$Dns = \sum_{i=1}^{n} (D_i^{'})_w \tag{10}$$

5.2 Controller Inputs

The controller algorithm receives as input the following variables.

5.2.1 Weight: The units of electricity required to meet the power demand consumption of each device from the smart grid. Represented as $\{(D_1)_w, (D_2)_w, (D_3)_w, \cdots, (D_n)_w\}$ for schedulable devices and $\{(D_1^{'})_w, (D_2^{'})_w, (D_3^{'})_w, \cdots (D_n^{'})_w\}$ for non-schedulable devices. $(D_i)_w$ and $(D_i^{'})_w$ stand for the weight/demand of the i^{th} schedulable and non-schedulable device.

5.2.2 Priority: It signifies the importance of any device among the plethora of home appliances for consideration while scheduling decisions are made. The assumption made in our work is priority and importance are inversely related. Represented as $<(D_1)_p, (D_2)_p, (D_3)_p, \cdots, (D_n)_p>$ for schedulable devices while $<(D_1^{'})_p, (D_2^{'})_p, (D_3^{'})_p, \cdots, (D_n^{'})_p>$ for non-schedulable devices.

$$priority \quad \alpha \quad \frac{1}{importance} \tag{11}$$

$(D_i)_p$ and $(D_i^{'})_p$ stand for the priority of i^{th} schedulable and non-schedulable device.

5.2.3 Peak: It signifies the cut-off load for the smooth operations of the smart grid. It helps to set the *threshold* value beyond which the smart grid becomes prone to errors. We have considered the maximum permissible value the grid can sustain without fault is 10% less than the given peak value.

5.2.4 Device Queues: They signify the allied data structure for maintenance of the smart device information. Each smart device along with its associated information is stored in the form of a linked list. Separate device queues i.e. array of m and k devices are maintained for schedulable and non-schedulable devices. The smart devices unscheduled at any point of time, becomes part of device queue $(D)_{left}$. The optimal order of scheduling of the smart devices $i.e.$ schedulable and unschedulable is maintained in the device queue denoted as $(D)^{opt}$. The queues $(D_i)^m_{i\leftarrow 1}$ and $(D_i^{'})^k_{i\leftarrow 1}$ specify the list of all schedulable and unschedulable devices while $(D_i)_{left}$ and $(D_i^{'})_{left}$ represent the list of unscheduled schedulable and unschedulable devices respectively. The collection of schedulable devices selected for smart scheduling are updated in the queue $(D_i)_{selected}$, for unschedulable devices in $(D_i^{'})_{selected}$.

5.3 Overview of SmartPeak

The proposed smart-peak algorithm aims to achieve the optimal scheduling of the schedulable devices across a finite time duration of a day for maintenance of customers load profile below the peak level whilst attaining maximum satisfaction level of the customer. In this paper, the satisfaction level of the customer is measured by the amount of electrical units reduced in comparison to potential electrical units consumption by virtue of scheduling of smart appliances. Initially the satisfaction is set to 100% i.e. no reduction is allowed and the load is set to 90 % of the peak level. Lines 3-7 of algorithm 1 represents the initialization section. Initialization pf optimal order device scheduling queue $(D)^{opt}$ along with selected devices queue $(D_i)_{selected}$ are initialized to *null*. The queues of devices waiting to be scheduled for schedulable $(D_i)_{left}$ and non-schedulable $(D_i^{'})_{left}$ are initialized with all the schedulable and unschedulable devices initially while their combination is updated into a common unscheduled devices queue $(D)_{left}$. The gist of SmartPeak algorithm is allocated to lines 8-27. An iterative approach is followed until devices are left which have not been used at least once. The aggregated load of the unschedulable devices is checked against the estimated *load* of the grid. If the aggregated demand is less than the estimated load, then all the unschedulable devices are switched on and they are simultaneously removed from the list of devices waiting in queue $(D)_{left}$ to be scheduled as shown in lines 10-12. While there are still devices left to be scheduled in $(D)_{left}$, sorting of the devices is done using the *knapsack* approach such that devices with highest priority and less weight can be scheduled earlier. A cut-off condition and selection of devices for scheduling is listed in lines 16-21. The lines 19 and 21 are self explanatory of updation of device queues $(D_i)_{selected}$ and $(D)^{opt}$ respectively. Lines 22-29 represent cases when the specified *load* is less than *Dns*, then sorting of the unschedulable device queue is done in increasing order of their weights. Subsequently unschedulable devices are selected and turned on one-by-one until the aggregated sum of the weights of the unschedulable devices does not cross the estimated load of the grid. The optimal schedule of devices is finally updated in the device queue $(D)^{opt}$.

5.4 SmartPeak Vs Dynamic Graph Modeling

Scheduling of devices(schedulable and non-schedulable) in a given time slot using sorting generally depends upon two prime factors: weight and priority. The specified weight of each smart appliance is the electricity units demand it applies for from the smart grid. Priority refers to the individual quota in a ranking-based system of a smart appliance to be considered during scheduling. Using the equation 11 the usage of the device is defined as:

$$usage\ of\ device \quad \alpha \quad \frac{1}{(weight\ of device * priority\ of\ device)} \tag{12}$$

$$= \frac{k}{W * P}$$

where k is a constant. Scheduling according to sorted order using equation 11 and 12 fails to capture important aspects like start and end running time of appliances, threshold values of electricity consumption per-appliance basis etc. SmartPeak takes as input the individual weights of devices, threshold load, various device queues to come up with a sophisticated approach for scheduling of

Algorithm 1: Iterative Algorithm for Load-Comfort Trade-off

Input : $Ds, Dns, peak, n, Device-Queues$
Output: $(D)^{opt}$

1 $load \leftarrow 0.9 * peak$
2 $count \leftarrow 1$
3 $(D_i)_{left} \leftarrow (D_i)^m_{i\leftarrow 1}$
4 $(D_i^{'})_{left} \leftarrow (D_i)^k_{i\leftarrow 1}$
5 $(D)_{left} \leftarrow (D_i)_{left} \cup (D_i^{'})_{left}$
6 $(D)^{opt} \leftarrow \phi, (D_i)_{selected} \leftarrow \phi, (D_i^{'})_{selected} \leftarrow \phi$
7 **while** $(D)_{left} \neq \phi$ && $count \neq n$ **do**
8     $count \leftarrow count + 1$
9   **if** $Dns < load$ **then**
10         $(D_i^{'})^k_{i\leftarrow 1} = on$
11         $(D)_{left} \leftarrow (D)_{left} - (D_i^{'})^k_{i=1}$
12   **while** $(D_i)_{left} \neq \phi$ **do**
13             $select\ and\ schedule\ D_i\ from\ (D_i)^m_{i\leftarrow 1}\ using$
14             $Knapsack\ algorithm$
15   **if** $(Dns + \sum_{j\leftarrow 1}^{i}((D_j)_{selected})_w > load)$ **then**
16   **break**
17             $(D_i)_{left} \leftarrow (D_i)_{left} - 1$
18             $(D_i)_{selected} \leftarrow (D_i)_{selected} \cup (D_i)$
19   **end**
20         $(D)^{opt} \leftarrow (D_i^{'})^k_{i\leftarrow 1} \cup (D_i)_{selected}$
21   **else**
22         $sort\ D_i^{'}\ in\ (D_i^{'})_{left}\ such\ that:\ (D_i^{'})_w < (D_{i+1}^{'})_w$
23         $schedule\ D_i^{'}\ such\ that:$
24                 $\sum_{i=1}^{t}(D_i^{'})_w < load$
25         $using\ lightest\ job\ first\ scheduling\ and\ t <= k$
26         $(D_i^{'})_{left} \leftarrow (D_i^{'})_{left} - 1$
27         $(D_i^{'})_{selected} \leftarrow (D_i^{'})_{selected} \cup (D_i^{'})$
28         $(D)^{opt} \leftarrow (D_i^{'})_{selected} \cup (D)^{opt}$
29   **end**
30 **end**

device(s). Despite being slightly more complex than dynamic graph modeling, SmartPeak can still yield better promising results.

5.5 Complexity Analysis

In this section, we will look at the overall time and space complexity of the proposed SmartPeak algorithm. If k represents the number of unschedulable devices, it takes $O(k)$ time presented to switch on all the unschedulable devices. If m denotes the number of schedulable devices and w represents weight of a $knapsack$, then sorting and scheduling schedulable devices can be achieved in $O(m * w)$ time. sorting of the unschedulable devices in increasing order of their weights can be achieved in $O(k * \log k)$ units of time. The units of time required for calculation of aggregate load demand of k unschedulable devices and m schedulable devices are $O(k)$ and $O(m)$ respectively. The overall time complexity for scheduling the devices i.e. schedulable or unschedulable is represented as follows:

$$Complexity = O(k) + O(m) + O(k) + O(m * w) + O(k * \log k) \tag{13}$$

If an equal number of schedulable and unschedulable devices are needed to be scheduled, then the equation 13 can be represented as:

$$Complexity = O(n) + O(n) + O(n) + O(n * w) + O(n * \log n) = O(n * \log n + n * w) \tag{14}$$

If T_s represents the minimum number of iterations of the proposed SmartPeak algorithm, then the conditional statement checking until any device is left unscheduled at least once, the overall complexity updates to:

$$Complexity = T_s * O(n * \log n) \tag{15}$$

If $EndTime$ and $StartTime$ represent the ending and starting timestamp of running of a smart appliance, the expected number of iterations $E(T_s)$, the SmartPeak algorithm is expected to run is represented below:

$$E(T_s) = (EndTime - StartTime)/(TimeSlotLength) \tag{16}$$

So, the overall time complexity of the proposed SmartPeak algorithm computes to:

$$\begin{aligned} T(n) &= O(n * \log n + n * w) * ((EndTime - StartTime)/ \\ &\qquad (TimeSlotLength)) \\ &\qquad or \\ &= O(n * (\log n + load - Dns)) * ((EndTime - StartTime)/ \\ &\qquad (TimeSlotLength)) \end{aligned} \tag{17}$$

where StartTime and EndTime represent the duration of scheduling of the device and TimeSlotLength represents the considered time period according to user convenience (NightHours:- 00:00 - 06:30, 21:00 - 24:00, MorningHours:- 06:30 - 11:30, DayHours:- 11:30 - 17:00, EveningHours:- 17:00 - 22:00). The overall space complexity of the proposed algorithm is $O(n * l)$ where l represents the number of independent information we store for each device irrespective of type.

6 CONCLUSION

Availability of granular level (in horizons of seconds or minutes) electricity consumption data in smart urban households has contributed effectively for better configuration of demand side management and demand response techniques in a smart grid environment. In this paper, we propose usage of machine learning techniques for forecasting overall energy consumption along with SmartPeak algorithm for smart building equipped with intelligent appliances. The proposed approach accounts for environmental explanatory variables of outside air temperature and ambient light. The developed procedure does a performance comparison of two popular machine learning regression techniques $i.e.$ MLR and SVR using RMSE as a performance metric for estimating consumption of an entire building. The results obtained in a case study demonstrated SVR outperforms MLR in terms of consumption forecasting. Additionally the proposed SmartPeak algorithm schedules the smart appliances $i.e.$ schedulable and non-schedulable devices considering the overall priorities of appliances and the threshold load of the building.

7 REFERENCES

[1] Joseph Iwaro and Abraham Mwasha. "A review of building energy regulation and policy for energy conservation in developing countries". In: *Energy Policy* 38.12 (2010), pp. 7744–7755.

[2] Hyunjoo Kim, Annette Stumpf, and Wooyoung Kim. "Analysis of an energy efficient building design through data mining approach". In: *Automation in construction* 20.1 (2011), pp. 37–43.

[3] Fu Xiao and Cheng Fan. "Data mining in building automation system for improving building operational performance". In: *Energy and buildings* 75 (2014), pp. 109–118.

[4] Omar Y Al-Jarrah et al. "Efficient machine learning for big data: A review". In: *Big Data Research* 2.3 (2015), pp. 87–93.

[5] Cheng Fan, Fu Xiao, and Chengchu Yan. "A framework for knowledge discovery in massive building automation data and its application in building diagnostics". In: *Automation in Construction* 50 (2015), pp. 81–90.

[6] Jie Zhao et al. "Occupant behavior and schedule prediction based on office appliance energy consumption data mining". In: *CISBAT 2013 Conference-Clean Technology for Smart Cities and Buildings*. 2013, pp. 549–554.

[7] Kaile Zhou and Shanlin Yang. "Understanding household energy consumption behavior: The contribution of energy big data analytics". In: *Renewable and Sustainable Energy Reviews* 56 (2016), pp. 810–819.

[8] Zhun Yu. "Mining hidden knowledge from measured data for improving building energy performance". PhD thesis. Concordia University, 2012.

[9] Tobore Ekwevugbe et al. "Improved occupancy monitoring in non-domestic buildings". In: *Sustainable cities and society* 30 (2017), pp. 97–107.

[10] Dimosthenis Ioannidis et al. "Occupancy driven building performance assessment". In: *Journal of Innovation in Digital Ecosystems* 3.2 (2016), pp. 57–69.

[11] Rajib Rana et al. "Novel activity classification and occupancy estimation methods for intelligent HVAC (heating, ventilation and air conditioning) systems". In: *Energy* 93 (2015), pp. 245–255.

[12] Dennis Atabay et al. “Self-adapting building models and optimized hvac scheduling for demand side management”. In: (2013).

[13] Alfonso Capasso et al. “A bottom-up approach to residential load modeling”. In: *IEEE Transactions on Power Systems* 9.2 (1994), pp. 957–964.

[14] JS Haberl and Sabaratnan Thamilseran. “The great energy predictor shootout II”. In: *ASHRAE journal* 40.1 (1998), p. 49.

[15] Geoffrey KF Tso and Kelvin KW Yau. “Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks”. In: *Energy* 32.9 (2007), pp. 1761–1768.

[16] Bing Dong, Cheng Cao, and Siew Eang Lee. “Applying support vector machines to predict building energy consumption in tropical region”. In: *Energy and Buildings* 37.5 (2005), pp. 545–553.

[17] Yangyang Fu et al. “Using support vector machine to predict next day electricity load of public buildings with sub-metering devices”. In: *Procedia Engineering* 121 (2015), pp. 1016–1022.

[18] Daniel Holcomb, Wenchao Li, and Sanjit A Seshia. “Algorithms for green buildings: Learning-based techniques for energy prediction and fault diagnosis”. In: *Google Scholar, UCB/EECS-2009-138* (2009).

[19] James A Rodger. “A fuzzy nearest neighbor neural network statistical model for predicting demand for natural gas and energy cost savings in public buildings”. In: *Expert Systems with Applications* 41.4 (2014), pp. 1813–1829.

[20] Nastaran Bassamzadeh and Roger Ghanem. “Multiscale stochastic prediction of electricity demand in smart grids using Bayesian networks”. In: *Applied energy* 193 (2017), pp. 369–380.

[21] Christopher O Adika and Lingfeng Wang. “Autonomous appliance scheduling for household energy management”. In: *IEEE transactions on smart grid* 5.2 (2014), pp. 673–682.

[22] Xiaodao Chen, Tongquan Wei, and Shiyan Hu. “Uncertainty-aware household appliance scheduling considering dynamic electricity pricing in smart home”. In: *IEEE Transactions on Smart Grid* 4.2 (2013), pp. 932–941.

[23] Amir-Hamed Mohsenian-Rad et al. “Autonomous demand-side management based on game-theoretic energy consumption scheduling for the future smart grid”. In: *IEEE transactions on Smart Grid* 1.3 (2010), pp. 320–331.

[24] Michael Angelo A Pedrasa, Ted D Spooner, and Iain F MacGill. “Coordinated scheduling of residential distributed energy resources to optimize smart home energy services”. In: *IEEE Transactions on Smart Grid* 1.2 (2010), pp. 134–143.

[25] Raj Mani Shukla, Prasanna Kansakar, and Arslan Munir. “A Neural Network-based Appliance Scheduling Methodology for Smart Homes and Buildings with Multiple Power Sources”. In: *Nanoelectronic and Information Systems (iNIS), 2016 IEEE International Symposium on.* IEEE. 2016, pp. 166–171.

[26] Bhagya Nathali Silva, Murad Khan, and Kijun Han. “Load Balancing Integrated Least Slack Time-Based Appliance Scheduling for Smart Home Energy Management”. In: *Sensors* 18.3 (2018), p. 685.

[27] Michael J Crawley. *The R book Second edition.* 2013.

[28] VN Vapnik. “Estimation of dependences based on empirical data. 1982”. In: *NY: Springer-Verlag* (1995).

[29] V Vapnik. “The nature of statistical learning theory Springer New York Google Scholar”. In: (1995).

[30] Yixuan Wei et al. “A review of data-driven approaches for prediction and classification of building energy consumption”. In: *Renewable and Sustainable Energy Reviews* 82 (2018), pp. 1027–1047.

[31] Nipun Batra et al. “It’s Different: Insights into home energy consumption in India”. In: *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings.* ACM. 2013, pp. 1–8.