2018 17th IEEE International Conference on Machine Learning and Applications

Forecasting Residential Energy Consumption: Single Household Perspective

Xiaoou Monica Zhang*, Katarina Grolinger*, Miriam A. M. Capretz*, Luke Seewald§
*Department of Electrical and Computer Engineering
Western University, London, Ontario, Canada, N6A 5B9
Email: {xzha28, kgroling, mcapretz}@uwo.ca,
§London Hydro, London, Ontario, Canada, N6A 4H6
Email: seewaldl@londonhydro.com

***Abstract*—With the development of smart electricity metering technologies, huge amounts of consumption data can be retrieved on a daily and hourly basis. Energy consumption forecasting facilitates electricity demand management and utilities load planning. Most studies have been focussed on commercial customers or residential building-level energy consumption, or have used behavioral and occupancy sensor data to characterize an individual household's electrical consumption. This study has analyzed energy consumption at single household level using smart meter data to improve residential energy services and gain insights into planning demand response programs. Electricity consumption for anonymous individual households has been predicted using a Support Vector Regression (SVR) modelling with both daily and hourly data granularity. The electricity usage data set for 2014 to 2016 was obtained from a Canadian utility company. Exploratory data analysis (EDA) was used for data visualization and feature selection. The analysis presented here demonstrates that forecasting residential energy consumption for individual households is feasible, but the accuracy is highly dependable on household behaviour variability.**

I. Introduction

Canadian households consumed 1.4 million Tera-joules of energy in their homes in 2013, up 7.2% from 2011. Electricity accounts for 44.6% of the total energy consumed by those Canadian residential customers [1]. The accelerated development of smart metering technologies and the Green Button initiative [2] has made it possible to measure, collect and present electricity consumption information for residential customers. Use of Home Area Networks (HAN) and Demand Response (DR) have brought a new focus on individual households [3]. Residential households voluntarily participate in DR programs, but the energy planning and estimation for these single households before the program can hardly be made without single household level energy prediction. Some utilities are issuing usage predictions or billing forecasts for residential customers to provide as an innovative service, helping households plan their consumption and reduce energy bills. Therefore, the potential of residential energy consumption prediction has gradually been recognized by governments and research institutes. Modelling and forecasting household electricity consumption using smart meter data helps utilities to implement effective load management on the demand side and to provide personalized residential services [4].

Electricity usage at the individual household level shows high variance because it depends on users' lifestyle, occupancy behavior, building characteristics, weather, and calendar information [5–7]. Feeding machine learning models with the data from energy smart meters and other relevant factors to infer the energy consumption for future days or hours is known as sensor-based forecasting approach [8]. Previous research [9–12] has established the prediction accuracy of sensor-based approaches for business or residential forecasting at the building level. In addition, these studies have investigated which machine learning techniques perform well for modelling commercial consumption. However, unlike the regularity seen in the workplace, with aggregated electric consumptions on routine schedules, more irregularity is foreseen in residential electrical consumption. Most households exhibit low base consumption on a daily basis, and load profiles for appliances such as air conditioners, clothes washers and dryers, pool pumps, and electric heaters are highly dynamic [13].

Energy management studies have been carried out on residential buildings [8, 10, 14] with datasets from multi-family residential buildings at an aggregated level or with sensor data collected to represent user behaviors and occupancy. Nevertheless, the energy consumption pattern of a single household can bring insights into residential behaviors, and enable the utility companies to provide personalized service for the residential market. Thus the prediction research has current and practical value to energy consumers. Meanwhile, installing various sensors to capture home users' behaviors and occupancy is not practical in most cases. Surveys to acquire detailed customer profiles would be expensive, impractical, time-consuming, and with low customer participation [15]. Forecasting households' electrical consumption using the raw data collected from commonly deployed smart meters in their home with weather variables [16, 17] and calendar information [14, 18, 19] is a cost-effective approach to gain insights into residential customers energy consumption and optimize energy efficiency programs.

This work has explored residential energy usage on a dataset representing fifteen randomly selected residential households with 534,966 lines of meter readings from a local utility company. The dataset contains three years (2014-2016) of hourly electricity consumption data from anonymous house-

978-1-5386-6805-4/18/$31.00 ©2018 IEEE
DOI 10.1109/ICMLA.2018.00024

holds' smart meter readings, without any known dwelling properties, occupation, nor household's socio-economic status. This study has revealed the capabilities of the support vector regression model, which is one of the most popular machine learning approaches for business energy prediction [18], in the field of residential energy consumption. Due to the variability of individual behaviors, prediction accuracy varies among families.

Traditional utility prices involve a set rate per kilowatt-hour that fluctuates between summer and winter [20]. A time-of-use (TOU) rate plan is a sliding-rate scale structured according to on-peak, mid-peak, and off-peak times of day [20]. In Ontario, TOU rates are defined by the Ontario Energy board [21], and are mandatory for residential customers across the province. In addition to energy forecasting, this study also examines the impact of TOU pricing schemes on residential electrical usage.

The remaining sections of this paper are organized as follows. Section II gives an overview of related studies; Section III introduces Support Vector Regression (SVR); Section IV describes the machine learning methodology; Section V presents the implementation, experimental results and an evaluation of the algorithm's correctness; and Section VI presents conclusions.

II. Related Work

The number of research studies in energy consumption, including annual consumption for various uses, characteristics impacting energy usage and consumption prediction, is dramatically increasing with the deployment of smart meters and other data collection methods. Many studies have explored machine learning approaches for modelling electrical consumption, applied in both commercial and residential sectors.

To reduce the impact of individual user's stochasticity in energy consumption, most energy modelling studies work with aggregated data in either residential buildings [7, 10, 12, 17, 18] or commercial buildings [8, 9, 11]. For example, Jain et al. presented energy modelling results on multi-family residential buildings [10], with single-families' consumption aggregated at building level. They built energy forecasting models using SVR with various data granularities. The results indicated that the most effective models were built with hourly consumption at floor level and had a standard error (standard deviation of the results obtained using the bootstrapping resampling method) of 28%. Liu et al. proposed a forecasting engine [12] based on a sliding window, empirical mode decomposition (SWEMD) and an Elman neural network (IENN) to predict the electricity load at the building level. The research of Grolinger et al. [9] demonstrated that both neural networks (NN) and support vector machine (SVM) are accurate in the consumption prediction for event-organizing venues using energy consumption data and event-related attributes. Zhang et al. investigated an institutional building's energy consumption using a weighted SVR [22], which was used to forecast half-hourly and daily electrical consumption. The work of Kaytez et al. [23] evaluated different algorithms and concluded that the result of the proposed Least Squares Support Vector Machine (LS-SVM) model provided a quick prediction with higher accuracy than both traditional regression analysis and Artificial Neural Networks (ANN).

In the area of individual residential energy forecasting, research studies have been carried out on highly detailed datasets [14, 24], including demographic survey information, dwelling characteristics, appliance ownership, and occupancy detection, to predict household's electrical loads. These are typical bottom-up approaches [25]. A study focusing on residential buildings by Edwards et al. [14] considered an extensive research data set: three homes with 140 sensors collecting human behaviors such as opening/closing refrigerators and using ovens, as well as occupancy patterns. They presented results for commercial consumption prediction and residential consumption prediction and concluded that NN-based methods are the best methods for commercial consumption prediction. The prediction accuracy comes from decomposition of electrical usage behaviors by a large number of sensors in the experiments, and can hardly be replicated for large-scale and cost-sensitive field application. Li and Dong built an occupancy prediction model [26] based on a Markov model using data obtained from occupancy sensors in four residential houses.

Without deploying behavioral sensors, other studies have focussed on behavior pattern clustering and stochastic simulation approaches [7, 15, 17] to recognize and simulate occupant behavioral patterns with clustering algorithms.

As for feature selection, Beckel et al. identified four main categories of factors that dramatically affect electricity consumption: weather and location, dwelling characteristics, appliance and electronics inventory, and occupancy and behavior [15]. Taking a more practical and feasible approach, recent studies have looked into ways to use temporal variables, such as calendar information [11] and weather variables and forecasts, which can be obtained from local or regional weather stations [9].

A wide range of modelling techniques have been used to predict electricity load, including ANN [12, 14], support vector machines (SVMs) [14, 22, 26], autoregressive integrated moving average (ARIMA) models [27], regressions models [27], clustering techniques [15] and empirical mode decomposition (EMD) [12]. NN have been extensively used for industrial electricity forecasting, whereas SVR has been successfully used to solve nonlinear regression and time series problems [22]. SVR performed best in some residential energy experiments [14].

In this study, SVR is used to predict fifteen households' residential electricity consumption. These households are anonymous residents of London, Ontario, with a home-installed smart meter for electricity measurement, but without any additional sensing devices.

III. Background

This section introduces SVR and performance measures.

A. Support Vector Regression

Support vector machines (SVM) [28] are supervised machine learning models used for classification problems. SVMs essentially consist of an optional kernel and optimizer algorithm; kernel transforms non-linear data into high-dimensional space and makes the data linearly separable, whereas optimizer is responsible for finding a decision boundary. Because SVM seeks to minimize an upper bound of the generalization error consisting of the sum of the training error and a confidence level, SVM usually achieves higher generalization performance than other machine learning techniques. SVR is a version of SVM for regression problems: it draws a decision boundary (or a hyperplane) to solve function fitting problems. Fig. 1 shows a non-linear regression with an epsilon intensive band. Sometimes data sets are linearly non-separable and must be mapped onto an N-dimensional space, in which case an (N-1)-dimensional separating hyperplane must be found. However, the process is computationally expensive. A suitable kernel trick significantly reduces the computational cost.

The relationship between the inputs $x_1, x_2, ..., x_n$ and the output Y is determined as:

$$Y = W\varphi(x) + b \tag{1}$$

where $\varphi(x)$ is a kernel function. Coefficients W and b are determined by minimizing the following function:

$$\frac{1}{2}\|w\|^2 + C\frac{1}{N}\sum_{i=1}^{N}\xi_i + \xi_i^* \tag{2}$$

subject to the constraints:

$$y_i - w^T\varphi(x_i) - b \leq \epsilon + \xi_i$$
$$w^T\varphi(x_i) + b - y_i \leq \epsilon + \xi_i^*$$

In the above equations, w is a weight vector and C is the cost of making an error. ξ_i and ξ_i^* are the residuals beyond the ϵ boundary, as shown in Fig. 1.

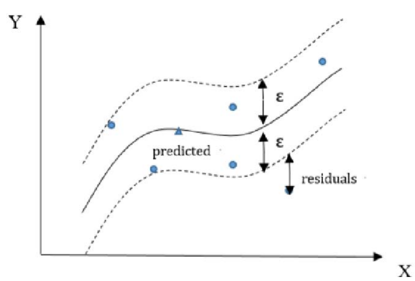


Fig. 1: Non-linear regression with epsilon intensive band.

B. Performance Measures

To determine the performance of the prediction model, this work uses the Mean Absolute Percentage of Error (MAPE) for validation.

The MAPE metric is a widely-used measure of prediction accuracy in electricity predictions studies [9, 22, 24]. It generally represents accuracy as a percentage, defined by the formula:

$$MAPE(\%) = \frac{100}{n}\sum_{j=1}^{n}\left|\frac{y_j - \hat{y_j}}{y_j}\right| \tag{3}$$

where y_j denotes the actual electricity consumption of household j, $\hat{y_j}$ denotes the predicted consumption, and n is the number of observations.

IV. Methodology

This work uses a machine learning approach, SVR modelling, for electricity forecasting. Both hourly and daily granularities are investigated, and prediction accuracy is evaluated. This section introduces the data preprocessing, exploratory data analysis and how the prediction model is built. Figure 2 illustrates the framework for individual household electrical forecasting.

A. Data preprocessing

The raw dataset from smart meters contains electricity consumption data measured in kilowatts-hours (kWh) for households on a time scale of one hour. Historical hourly weather and humidity data were obtained from the Canadian Government Official Website [29] as factors impacting residential energy consumption.

1) Data Cleansing: Some hourly data were missing and some were in wrong order in the meter readings. To avoid unexpected impacts on the prediction model, missing consumption cells were replaced by the average electrical consumption value of the previous and following hours, and the data were sorted chronologically. In the case of invalid/missing weather conditions or temperature/humidity data in the dataset downloaded from the government Web site, an average temperature/humidity value for the previous and following hours was calculated as a substitute for the missing hour. For an invalid weather condition (snow, cloudy, clear, etc.), the cell is filled using the last hour's weather condition.

2) Feature preparation: The cleaned and consistent dataset was reorganized, and additional new features were generated as follows:

- Weather Conditions: To simplify the impact from weather conditions, over 20 weather conditions in the original version were aggregated to seven main weather types: "Clear", "Cloudy", "Fog", "Haze", "Rain", "Snow", and "Ice", which represent different weather conditions in Canada in order of severity from low to high.
- Temperature: To predict electricity consumption, temperature is used because air-conditioning (AC) consumption accounts for a large proportion of all electricity consumption, especially in summer, as well as other residential electrical appliances.
- Humidity: Relative humidity data were retrieved from the Government of Canada Web site [29].
- Hour of the Day: 1 to 24. Hour of the day is an important attribute for forecasting energy consumption. Consumption at midnight is typically less than the consumption at

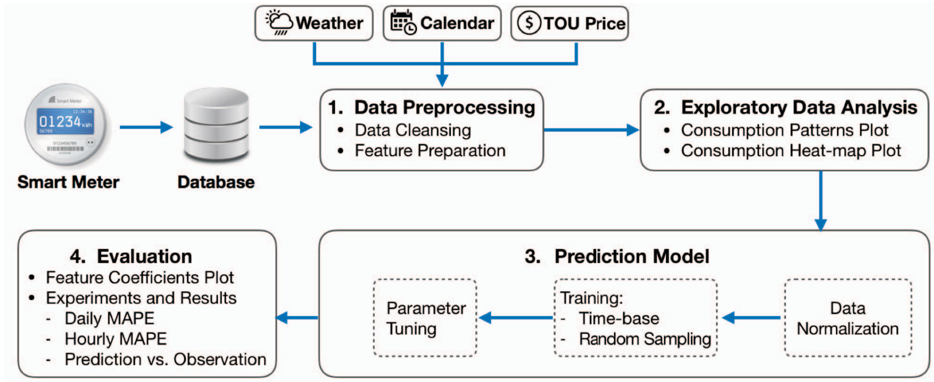


Fig. 2: Framework of individual household electricity forecasting.

7:00 pm, which was confirmed by an Integrated Energy Mapping Strategy [30].

- Day of the Week: 1 to 8 for weekdays, weekends and holidays. Monday through Sunday are represented by numbers 1 to 7, and 8 represents public holidays in Ontario.
- Week of the Year: 1 to 52 to represent the weeks of the year.
- Month: 1 to 12. Month was included as a factor to complement temperature and to reduce the probability of prediction errors.
- Season: Spring, summer, fall, and winter. Season was used as an attribute to detect the relationship between season and electricity usage and to improve forecasting accuracy.
- Price: The electricity TOU rate is defined by Ontario Energy board [21]. The on-peak, off-peak and mid-peak prices for each hour of the year are used as the pricing factor.

B. Exploratory data analysis

Compared with commercial electricity consumption, the most significant characteristic of residential electricity usage is that some families live regular lives, which can be reflected in their consumption patterns. However, other families are irregular in their energy usage. This leads to various consumption patterns among different families.

Investigating the correlations between the input features and energy consumption [17] makes it easier to choose the features that exhibit significant correlation with energy consumption for a particular household. This reduces the computing time required for parameters tuning by using a subset of variables significantly correlated to the consumption.

1) Consumption Patterns Plot: The stochasticity nature in family members' behaviors results in various electricity consumption patterns. Figure 3(a) shows two households' electricity usage trends and associated patterns from January 1, 2014, to December 31, 2016. In the figure, the x-axis denotes time, and the y-axis represents electricity consumption.

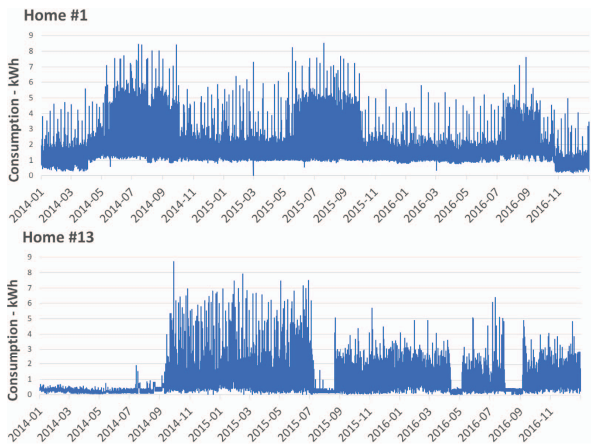
There is a visually detectable pattern for home #1. Its electricity usage ramps up during summer time and is relatively lower in winter. This is typical residential behavior in Canada due to the air-conditioning usage on high-temperature days, which boosts the electricity usage between May and October. For Winter heating system, gas is more widely used than electricity for residential homes in most Ontario houses and apartments because of its lower cost. Therefore, electricity usage drops from October to May of the next year. However, the electricity usage pattern is difficult to recognize for home #13. Less regularity in electricity consumption pattern can be deducted from the home #13 hourly trend.

2) Consumption Heat-map Plot: To obtain better data visualization, detailed 2-D consumption heat-maps were plotted for these two households, as shown in Fig. 3(b). In the figure, the x-axis denotes the hours in a week from Sunday to Saturday with a step of 6 hours, and the y-axis denotes the days from 2014 to 2016. Each hourly consumption figure is represented by a colored square according to a predefined "rainbow" color bar with colors from warm to cold. The more consumption that occurs in a particular hour, the warmer the color will be. For home #1, the periodical appearance of high usage during daytime in the summer months is evident, especially in the year 2014 and 2015. By contrast, less regularity can be identified from the consumption data visualization for home #13.

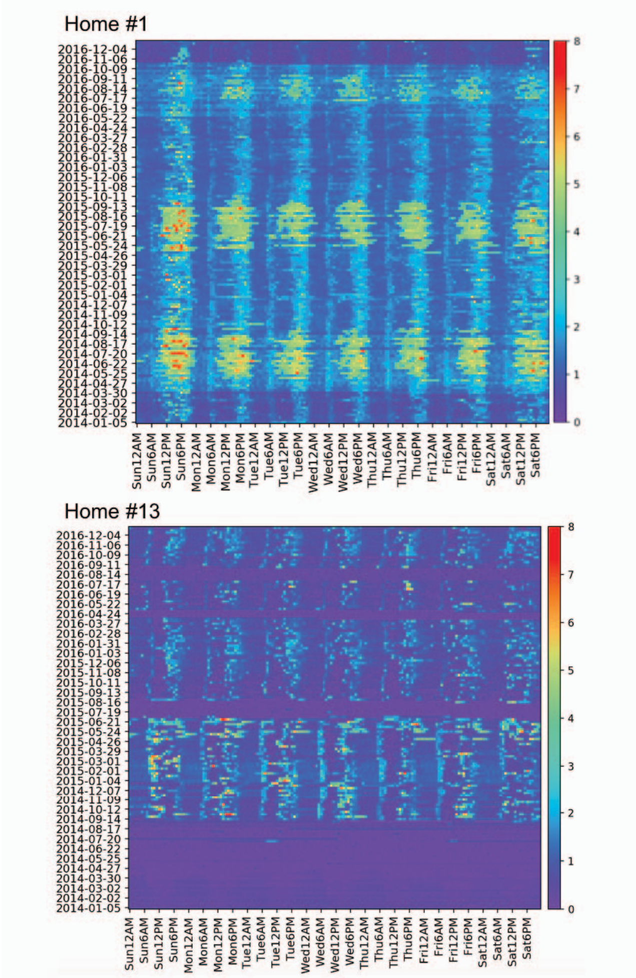
Residential characteristics including household's socio-economic status, its dwelling characteristics, employment status and even the number of persons living in the household are relevant to hour-by-hour residential electricity usage [15]. For example, in a rental residence with changing tenants and undetermined vacancy, it is difficult to predict electricity consumption.

C. Prediction Model

The prediction model was designed to work with SVR on both hourly and daily data granularities for every household. For hourly granularity, one observation was associated with one energy reading. Other features discussed in Section IV (B) were added to the energy reading data set. To explore the



(a) Three-year hourly consumption pattern.



(b) Three-year hourly consumption heat-map.

Fig. 3: Two households three-year hourly consumption.

accuracy at daily granularity, the hourly data were aggregated as follows to obtain the daily electricity consumption:

$$C_d = \sum_{i=1}^{24} C_{h_i} \tag{4}$$

where C_{h_i} is the electrical consumption for the i^{th} hour in a day.

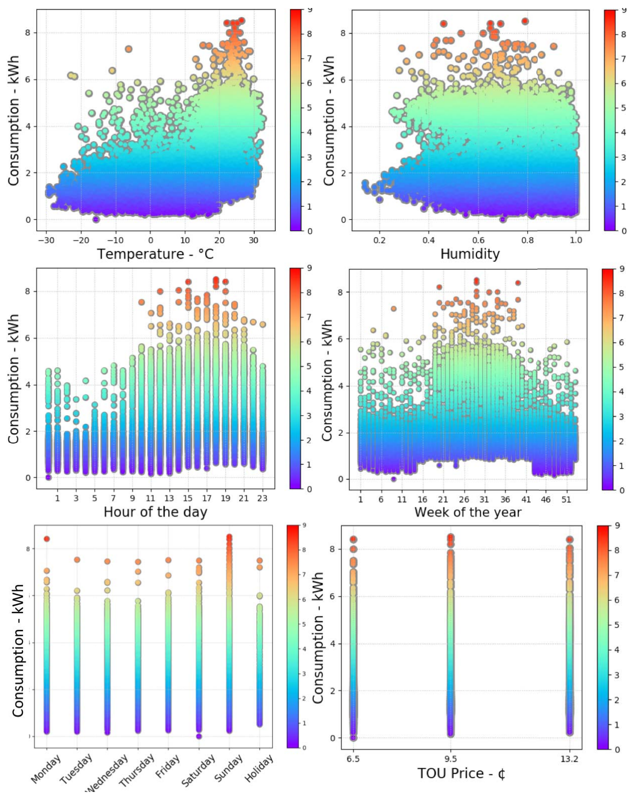
To train the prediction model, splitting time series into chronological sets has proved to be useful for predicting buildings energy consumption for businesses [9, 14]. However, irregularities and uncertainties over time for some residential customers may make a consecutive training set less accurate than random sampling. For the households, who have shown less regularity in hourly energy usage, applying random sampling to the data may lead to better accuracy. Hence, for every household, both consecutive time splitting and random sampling were used to predict residential electricity consumption. Random sampling may not be a realistic approach for energy forecasting and here it serves only as comparison and a way of exploring energy variability.

For the machine-learning regression model SVR, the parameters for the prediction model must be determined in the learning phase. With a nonlinear kernel function, two basics parameters should be determined in advance: the cost C, which denotes the penalty for errors greater than ϵ as shown in Fig. 1, and the nonlinear radial basis function (RBF) kernel coefficient γ. Combinations of model parameters constitute a model configuration, among which the best parameters must be chosen for best prediction accuracy. A grid search method was used to tune the model configuration. In this implementation, a set of C and γ parameters was constituted by assembling a grid of search parameters. Once the optimized parameters had been determined, the prediction error could be validated by model assessment. The SVR prediction module was implemented in the Python language.

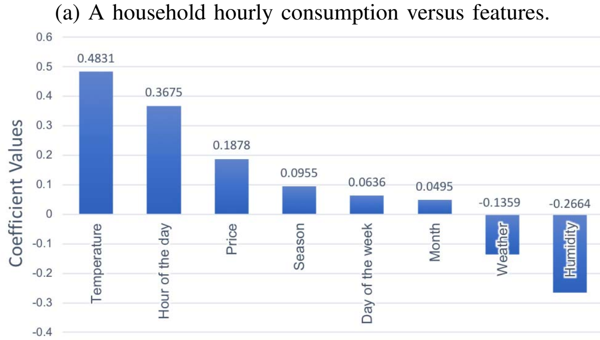
Because each household required SVR parameters tuning for both chronological splitting and random sampling at hourly and daily data granularities to obtain the best modelling configuration, there are $15 \times 2 \times 2 = 60$ passes of grid search tuning for the fifteen households on over 500,000 lines of data. A small cluster of servers was used, with each server having 24 Intel Xeon CPUs, 96 GB memory. To speed up the SVR parameters' cross-validation, a parallelized computing structure was implemented in Python to maximize the concurrently running processes.

V. Evaluation

The proposed method was tested using residential electricity consumption data provided by London Hydro, the electricity service provider in London, Ontario, Canada. Selecting fifteen anonymous households allows us to compare individual ones (as done in Fig 3) and at the same time captures diverse behaviors as demonstrated in the experiments. Exploratory analysis as described in Section IV was carried out for all fifteen households to visualize each household settled habits



(a) A household hourly consumption versus features.



(b) Pearson correlation coefficients for household features.

Fig. 4: Relationship between features and consumption.

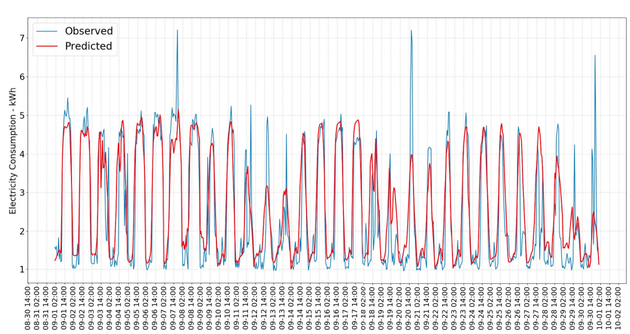


Fig. 5: A household hourly prediction vs actual observation.

and a daily routine. Then the SVR model was applied to each household respectively at different data granularities.

The dataset incorporated the hourly electricity consumption of fifteen households from 2014 to 2016. The original electricity consumption data contained 534,966 lines of hourly meter readings for fifteen households over three years. Time-based splitting was done by dividing data into two months intervals, with consecutive hourly readings in each part.

A. Feature Coefficients

A further exploratory analysis was performed for each household to analyze the relationship between important influencing factors and the household's electrical consumption.

Figure 4(a) visualizes the Pearson correlations between a household's three-year hourly consumption and features like "temperature", "humidity", "hour of the day", "week of the year", "day of the week", and "TOU price". "Temperature" has an obvious positive correlation with energy usage: the hourly consumption increases along with the rising temperature. By contrast, "humidity" does not have a positive correlation with consumption. As for "hour of the day", the hours from 2AM to 6AM accounted for the lowest electricity usage and the hours in the afternoon for higher consumption. On average, weeks 19 to week 39 in the year have higher hourly electrical consumption, corresponding to the summer and fall seasons. During a week, Sunday has the most energy usage compared with other days. Although we considered features "month" and "season", graphs for those features are not included as they show very similar patterns to the "week of the year" feature.

For the features used in the prediction, Pearson correlation coefficients with the hourly consumption of the same household were calculated, as shown in Fig. 4(b). "Temperature", "hour of the day" were the two most important factors with the top two high coefficients. Variables with negative coefficients have inverse relationships with hourly consumption for the households. The impact of "TOU Price" for the household comes after "Temperature", "hour of the day" and "Humidity", reflecting the family members' energy consumption habits. It is clear that households' correlation coefficients between features and consumption differ from each other. For example, a certain proportion of Ontario residents use electricity for heating during winter rather than gas, resulting in a remarkably high electrical consumption when the temperature is below freezing. Hence, non-positive correlations between temperature and consumption can be found for those families.

B. Experiments and results

Fig. 5 illustrates actual electricity usage and the predicted energy consumption values obtained by SVR model for one month (e.g. September 2015) for home#1 mentioned in section IV. Note that the prediction curve follows the fluctuation of the actual electricity usage during the day and night on an hourly basis, except for several peak usage hours. This occurs because of random variations due to individual home users.

The evaluation metrics for all fifteen households are shown in Table I. Although the consumption dataset for these fifteen households was a random sample from London's residential customers' energy data, it represented several categories classified by families' regularities in electrical consumption, as indicates by the column of "Accuracy Category" column. Most households had acceptable daily consumption MAPE, but the accuracy for hourly consumption fluctuates dramatically. The first two households (home#1 and #2) had both hourly and daily MAPE under 25 and showed regularities in their electrical consumption over time. These two families also had the highest hourly average consumptions (Fig. 6) over the three years. The households with ID from #3 to #9 had lower hourly prediction accuracy, but showed greater regularity if their hourly usage was aggregated into daily consumption. Those households had similarities in their consumption patterns over time, and gained better accuracy through time-based training and testing subsets splitting. By contrast, home #10 to #15 demonstrated less regularity over a continuous time period, and could be better forecasted using random sample splitting.

TABLE I: SVR modelling evaluation results and categories.

Home No.	**MAPE**		**Accuracy Category**	**Data Splitting Method**
	Hourly	**Daily**		
#1	23.31	12.78	Good hourly and daily accuracy	Time-based splitting
#2	24.42	14.46		
#3	35.82	17.30	Weak hourly, better daily accuracy	
#4	36.65	19.66		
#5	69.17	13.72	Poor hourly, better daily accuracy	
#6	43.41	21.89		
#7	61.07	21.31		
#8	53.81	25.61		
#9	67.96	24.14		
#10	36.05	22.55	Weak hourly, better daily accuracy	Randomly sampling
#11	44.63	22.66		
#12	33.33	24.8		
#13	45.33	29.31		
#14	40.34	34.49	Poor hourly and daily accuracy	
#15	64.38	34.95		

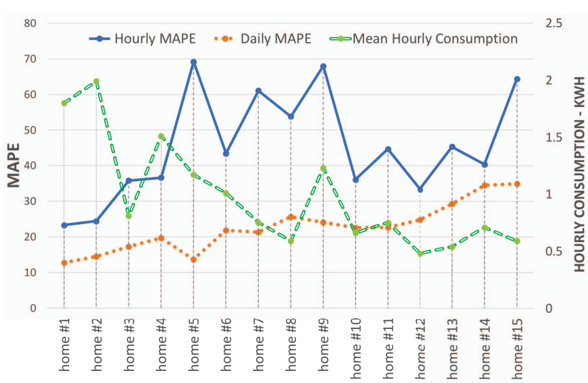


Fig. 6: Consumption MAPE and mean hourly consumption for the fifteen households.

TABLE II: Home #1 MAPE between all days and weekdays.

Home #1	**All Days**		**Weekdays**	
	Time-based Splitting	**Randomly Sampling**	**Time-based Splitting**	**Randomly Sampling**
MAPE	23.31	24.91	22.01	23.94

Higher variability in electricity consumption has been seen on weekends compared with weekdays. Table II shows an example of MAPE comparison between the whole dataset and the subset of weekdays for the household #1. The forecasting for weekdays exhibited lower MAPE errors for both time-based splitting and randomly sampling when processing the training and testing datasets. Figure 7 visualizes the weekday electricity consumption predictions for home #1 over four months. Seasonal changes from September to October had brought a substantial energy decrease, leading to an imperfectly matched prediction curve. However, for most hours, lower errors can be seen in the weekday usage forecasts from October to December 2015.

VI. Conclusions

This study explores the accuracy of machine learning SVR modelling approach in residential electricity consumption prediction. It confirmed the influence of various elements, including weather condition, temperature, humidity, holidays, hour of the day, day of the week, month, season, and electricity price on residential electricity consumption and provided an efficient and accurate approach to prediction. It also revealed, using the correlation coefficients, that hour of the day and temperature had the most significant impacts on electricity consumption in this case. Various methods were used to split the dataset into training and testing subsets. For households with similar electrical consumption over time, consecutive time-based splitting works better than randomly sampling the data. However, for those with irregular hourly electricity usage, sampling the whole dataset regardless of time and using 20% of the sample as a testing sub-dataset outperformed the time-based approach.

Because of the stochastic nature of single residential customers, daily data granularity achieved better prediction results than hourly granularity. Aggregating hourly consumption to daily is an effective way to mitigate the impact of randomness in the hourly behaviors of family members. The lowest single-household MAPE error was 12.78 for daily prediction and 23.31 for hourly prediction, and this further decreases to 22.01 (hourly) if only weekdays were counted for the same family.

Future work will extend to a large number of households and attempt to categorize them according to their energy consumptions, and explore the applicability of other machine-learning approaches.

Acknowledgment

This research has been partially supported by the NSERC CRD at Western University (CRD 477530-14). The authors

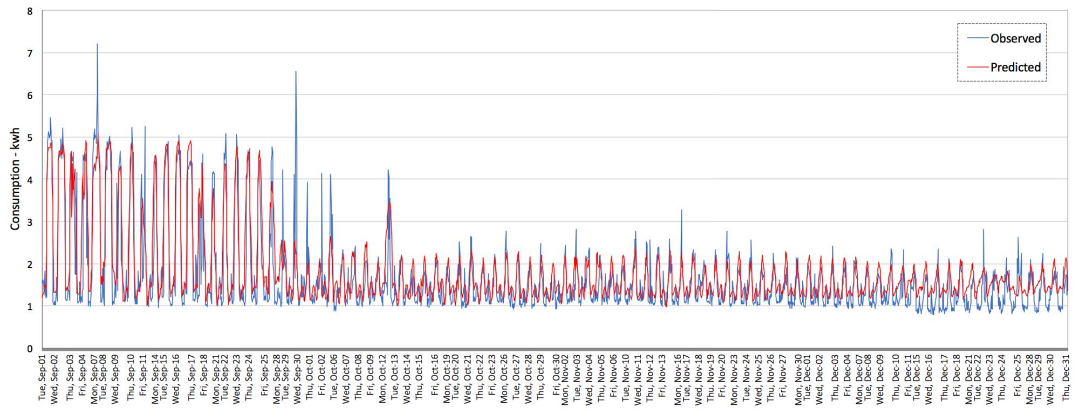


Fig. 7: A household weekday hourly prediction over four months.

would like to thank London Hydro for supplying industry knowledge and the Green Button platform.

References

[1] Statistics Canada, "Households and the environment survey: Energy use, 2013," http://www.statcan.gc.ca/daily-quotidien/160318/dq160318d-eng.htm, Mar 2016.
[2] US Department of Energy, "Green button," https://energy.gov/data/green-button.
[3] M. Pipattanasomporn, M. Kuzlu, and S. Rahman, "Demand response implementation in a home area network: A conceptual hardware architecture," *Proceedings of the IEEE Innovative Smart Grid Technologies (ISGT)*, pp. 1–8, 2012.
[4] Q. Sun, H. Li, Z. Ma, C. Wang, J. Campillo, Q. Zhang, F. Wallin, and J. Guo, "A comprehensive review of smart energy meters in intelligent energy networks," *IEEE Internet of Things Journal*, vol. 3, no. 4, pp. 464–479, 2016.
[5] A. Tascikaraoglu, A. Boynuegri, and M. Uzunoglu, "A demand side management strategy based on forecasting of residential renewable sources: A smart home system in turkey," *Energy and Buildings*, vol. 80, pp. 309–320, 2014.
[6] P. Lusis, K. R. Khalilpour, L. Andrew, and A. Liebman, "Short-term residential load forecasting: Impact of calendar effects and forecast granularity," *Applied Energy*, vol. 205, pp. 654–669, 2017.
[7] L. Diao, Y. Sun, Z. Chen, and J. Chen, "Modeling energy consumption in residential buildings: A bottom-up analysis based on occupant behavior pattern clustering and stochastic simulation," *Energy and Buildings*, vol. 147, pp. 47–66, 2017.
[8] H.-x. Zhao and F. Magoulès, "A review on the prediction of building energy consumption," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 6, pp. 3586–3592, 2012.
[9] K. Grolinger, A. L'Heureux, M. A. Capretz, and L. Seewald, "Energy forecasting for event venues: Big data and prediction accuracy," *Energy and Buildings*, vol. 112, pp. 222–233, 2016.
[10] R. K. Jain, K. M. Smith, P. J. Culligan, and J. E. Taylor, "Forecasting energy consumption of multi-family residential buildings using support vector regression: Investigating the impact of temporal and spatial monitoring granularity on performance accuracy," *Applied Energy*, vol. 123, pp. 168–178, 2014.
[11] S. S. Abdelkader, K. Grolinger, and M. A. Capretz, "Predicting energy demand peak using m5 model trees," *Proceedings of the IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 509–514, 2015.
[12] Y. Liu, W. Wang, and N. Ghadimi, "Electricity load forecasting by an improved forecast engine for building level consumers," *Energy*, vol. 139, pp. 18–30, 2017.
[13] B. Yildiz, J. Bilbao, J. Dore, and A. Sproul, "Recent advances in the analysis of residential electricity consumption and applications of smart meter data," *Applied Energy*, vol. 208, pp. 402–427, 2017.
[14] R. E. Edwards, J. New, and L. E. Parker, "Predicting future hourly residential electrical consumption: A machine learning case study," *Energy and Buildings*, vol. 49, pp. 591–603, 2012.
[15] C. Beckel, L. Sadamori, T. Staake, and S. Santini, "Revealing household characteristics from smart meter data," *Energy*, vol. 78, pp. 397–410, 2014.
[16] Y.-H. Hsiao, "Household electricity demand forecast based on context information and user daily schedule analysis from meter data," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 1, pp. 33–43, 2015.
[17] T. K. Wijaya, M. Vasirani, S. Humeau, and K. Aberer, "Cluster-based aggregate forecasting for residential electricity demand using smart meter data," *Proceedings of the IEEE International Conference on Big Data*, pp. 879–887, 2015.
[18] S. Humeau, T. K. Wijaya, M. Vasirani, and K. Aberer, "Electricity load forecasting for residential customers: Exploiting aggregation and correlation between households," *Proceedings of the IEEE Sustainable Internet and ICT for Sustainability (SustainIT)*, pp. 1–6, 2013.
[19] M. Rossi and D. Brunelli, "Electricity demand forecasting of single residential units," pp. 1–6, 2013.
[20] M. Hinterstocker, P. Schott, and S. von Roon, "Evaluation of the effects of time-of-use pricing for private households based on measured load data," *Proceedings of the IEEE International Conference on the European Energy Market (EEM)*, pp. 1–6, 2017.
[21] Ontario Energy Board, "Electricity rates," https://www.oeb.ca/rates-and-your-bill/electricity-rates.
[22] F. Zhang, C. Deb, S. E. Lee, J. Yang, and K. W. Shah, "Time series forecasting for building energy consumption using weighted support vector regression with differential evolution optimization technique," *Energy and Buildings*, vol. 126, pp. 94–103, 2016.
[23] F. Kaytez, M. C. Taplamacioglu, E. Cam, and F. Hardalac, "Forecasting electricity consumption: a comparison of regression analysis, neural networks and least squares support vector machines," *International Journal of Electrical Power & Energy Systems*, vol. 67, pp. 431–438, 2015.
[24] B. Dong, Z. Li, S. M. Rahman, and R. Vega, "A hybrid model approach for forecasting future residential electricity consumption," *Energy and Buildings*, vol. 117, pp. 341–351, 2016.
[25] A. Grandjean, J. Adnot, and G. Binet, "A review and an analysis of the residential electric load curve models," *Renewable and Sustainable Energy Reviews*, vol. 16, no. 9, pp. 6539–6565, 2012.
[26] Z. Li and B. Dong, "A new modeling approach for short-term prediction of occupancy in residential buildings," *Building and Environment*, vol. 121, pp. 277–290, 2017.
[27] K. Gajowniczek and T. Zabkowski, "Electricity forecasting on the individual household level enhanced based on activity patterns," *PloS one*, vol. 12, no. 4, p. e0174098, 2017.
[28] V. N. Vapnik and S. Kotz, *Estimation of dependences based on empirical data*. Springer-Verlag New York, 1982, vol. 40.
[29] Government of Canada, "Canada weather information," https://weather.gc.ca/.
[30] B. Gilmour, "City of london: Integrated energy mapping strategy (l-iems)," Canadian Urban Institute, Tech. Rep., 2011.