International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 1078-1091 www.ijaem.net ISSN: 2395-5252

An Implementation of Short-term Load Forecasting for Campus Microgrid based on Artificial Neural Network and Clustering method

Minh-Tuan Nguyen*, Ai-Linh Pham, Minh-Duc Ngo
Thai Nguyen university of Technology, Viet Nam
Corresponding Author: Minh-Tuan Nguyen

Date of Submission: 18-05-2024 Date of Acceptance: 28-05-2024

ABSTRACT: Recently, MicroGrid (MG) technology has been proposed as one of the most critical solutions for energy poverty problem. In MG technology, Micro-Grid Energy Management System (MG-EMS) plays an important role in control, monitoring and optimize the performance of the system. There are two general approaches: real-time immediate technique and predictive technique to meet the operational objectives such as cost, power sharing, and so on. In this paper, we investigate a newly developed method to predict the energy use in campus buildings based on Artificial Neural Network (ANN) in short-term time series from one hour up to one week. We propose a new method to analyze and extract the features from the historical data of load and temperature to generate the prediction of future energy consumption in the building based on Sparsified K-means. To evaluate the performance of the proposed approach, historical load data in hourly resolution collected from Chonnam National University in South Korea were used to train and test the system. The experimental results show that the proposed approach outperforms the conventional methods in term of features selection.

KEYWORDS: ANN, Campus building, Energy Management System, Electricity short-term load forecasting, Features selection, MicroGrid, Sparsified K-means.

I. INTRODUCTION

Short-term load forecasting (STLF) is basically aimed at predicting system load with a leading time of one hour up to one week, which is necessary for adequate scheduling and operation of power system.

STLF is one of the crucial components in energy management system, it plays an important role in providing the input data for load flow and contingency analysis. The more accuracy in load forecasting the more reliable the system will be.

Furthermore, deregulated electricity market is now becoming widespread, expanding from large-scale leverage to small-scale leverage in power system. Therefore STLF is even more important, not only for system operators, but also for market operators, and for any other market participants. In the typical micro-grid energy management system architecture, STLF is one of the core components in MG-EMS, which associated with other components such as optimization and data analysis components in scheduling and operating plans of energy transactions. In this new context, high forecasting accuracy and speed will result in decreasing operation cost and increasing revenue of system.

Recently, based on surveys of many researchers, buildings are one of the fastest growing consuming sectors and the demand for energy of buildings will continue to grow. It is estimated that the amount of energy consumed by buildings in 2015 was responsible for approximately 41% of the total energy consumption in European Union's and 73% of total electricity generated in the United States [1]-[2]. In Korea, according to the Korean energy consumption survey [3] which has been conducted every three years by the Korea Energy Economics Institute (KEEI), commercial building sector accounted for 30%-40% of total power consumption [4]. As a result, forecasting of building energy use has received much attention from many researchers.

However, building an accurate energy-forecasting model is a crucial task due to the influences of indoor factors such as electrical devices inside the building, the time range of building operation, the number of people working in the building and outdoor factors as the weather variables. There is a lot of valuable information from weather variables for forecasting task at a small micro-grids system [5].

Generally, the prediction model is essentially affected by the historical energy consumption factor. Each commercial building consists of various electrical loads such as the lighting system, air-condition, motors system and so on which are consistently used electrical power. In addition, other energy consuming devices such as computers, printers, fax machines and other office devices also contributed to the power consumption of a commercial building. Among them, air-condition is the most power consumption in commercial building at schools or universities. Moreover, the working time and the number of workers are other critical factors persuaded to the electrical load. The building with a higher number of workers using electrical devices consumes greater power load and conversely.

Moreover, the external weather conditions play an important role to produce the load-forecast model. There are numerous weather variables such as the temperature, dew point, dry bulb temperature, air pressure, humidity, wind speed and conditions of the cloud. Among them, the most impressive factor to the building load is the temperature which is able to change the indoor factor and the electrical power consumption.

In order to achieve the accurate forecasting results, all of the aforementioned factors should to be considered. However, based on the data collection system, the power consumption and the temperature are provided to build the predicted model.

A large number of load forecasting approaches have been proposed in various research papers, which can be categorized into three categories, which are conventional methods, artificial intelligence method and hybrid method. The conventional methods have been broadly adopted in load forecasting problem in the past. The conventional methods are capable of achieving satisfactory results when solving the linear problems. There are several kinds of the conventional approach based on time series models such as Auto-Regressive model (AR), Moving Average model (MA), Auto-Regressive Moving Average Model (ARMA), and Autoregressive Integrated Moving Average Model (ARIMA). The regression based approaches including linear regression (LR), Multiple Linear Regression (MLR), Regression Tree (RT) were broadly investigated. AR was applied in [7] based on the assumption that the current load is a linear combination with the previous load. Another advantage of using AR method is that AR method does not require pre-training. However, the results are only limit in few data. S.R Huang proposed the threshold AR models with the stratification rule to forecasting the hourly load demand on a power system [8]. In this paper, the optimal stratification rule attempts to remove any judgmental input and to render the threshold process entirely mechanistic. However, because the load data is non-stationary, it is extremely difficult to model the system.

ARMA is the combination of autoregressive with moving average method. The ARMA method is usually applied to stationary stochastic processes [9]. An adaptive ARMA model was adopted for short-term load forecasting of a power system in [10]. The authors proposed to use the available forecast error to update the ARMA method for the load forecasting.

ARIMA is an updated version of ARMA method. Due to some improvements, the ARIMA method is able to work with non-stationary stochastic processes. A real time load forecasting was introduced in [11].

On the other hand, the regression based approaches have been adopted in this research filed to forecast the electricity consumption. There are two main methods include of LR [12] and MLR [13]. The main advantages are that the changing between the dependent factors and independent factors can be determined via the regression based approaches. Regression tree was introduced for pattern-based forecasting time series task in [14].

However, the conventional methods may not be able to deal with non-linear factors such as the weather conditions and indoor factors which significantly impact on the prediction accuracy. Recently, Artificial Intelligence (AI) based methods have increased the most attention of various researchers, due to achieving high forecasting accuracy [16]. In [17], the back propagation neural network was applied to analyse the electrical power consumption of a residential building in China. The hourly temperature curve was modeled to forecast the electrical consumption by ANN [18]. Mai, Chung [19] applied the radial basis function of neural network to predict the load consumption in large building office. They investigated three different situations based on weather, namely, hot, cool and cold weather. Another research work based on ANN was exploited to forecast the electrical energy use in the cooling system [20]. In [15], the author proposed a new method to forecasting energy consumption in residential building via support vector regression method. Two crucial improvements based on procedure for generation of model inputs and subsequent model input selection using feature selection algorithms to the Support

Vector Regression (SVR) based load forecasting method are introduced in [21].

The hybrid method has been considered for load forecasting problem by the combination of the conventional method as ARIMA and AI method such as ANN, SVM. A hybrid approach that combines both ARIMA and ANN model was proposed to take the advantage of the unique strength of ARIMA and ANN models in linear and nonlinear modeling [22]. Another hybrid method based on ARIMA and SVM model was introduced in [23].

In this paper, a novel approach based on Sparsified K-means and Resilient back-propagation ANN algorithms for short-term load forecasting problem is proposed. The main contributions of this paper are summarized as follows.

1)The impact of time, temperature, humidity, wind speed to electric consumption is carefully analyzed. We conclude that the power load of campus building is highly affected by time and temperature.

2)The past 24 hours data including time calendar, temperature, and load are investigated to predict the electrical power at the current time. The past 24 hours data is extremely small data in comparison with other published researches which usually based on past one week data.

3)To reduce the impact of the variety of training data, Sparsified K-means algorithm is applied to classify the original training dataset to several groups which have same properties. In comparison with normal K-means approach, Sparsified K-means shows better performance on processing time, which is one of the most important requirements for real time prediction.

4)The Resilient back-propagation ANN is used to train the network. We randomly selected 70%, 20% and 10% of previous year data for training, validation and testing set to train the networks. The prediction performance is applied for the first week of the next year.

5)The performance of the proposed approach is evaluated. The experimental results show that MAPE, RMSE of the proposed method are smaller than other previous approaches. The strong correlation between the predicted and real value is depicted by R-squared value. In addition, the effects of time calendar and temperature to prediction results are analyzed.

This paper is structured as follows: Section 2 introduces in detail the statistical analysis of actual dataset using in this research work. Section 3 presents the case study on the load forecasting methodology in the acquisition dataset. In section 4, we examine the proposed method performance in comparison with the other proposed method. Section 5 concludes the present paper.

II. DATASET ANALYSIS

This study utilizes the dataset collected in two years from March 2014 to February 2016 for a university building at Chonnam National University, South Korea. The building consists of offices, classes and laboratories. The energy use, temperature and wind speed are collected at every one hour interval. However, the humidity is collected after every three hours. As mention previously, the prediction accuracy depends on the indoor data as previous load consumption and the outdoor data such as the temperature, humidity, and wind speed. Each factor will be analyzed in details in the following sections.

2.1 Load dataset

The original load profile recorded by the smart metering system has two critical anomalies due to improperly calibrated values of devices. The first is the missing data in some hours in one day or in all days. The missing data occurs because no data is stored. The loss of data at some point significant effect on the system performance. To deal with this problem, linear interpolation method was adopted due to the simple and fast calculation. The second is the presence of outliers' data. The outlier consists of zero, minus and suddenly changing in one or two hours without any reason. The outliner values are imputed via the average between the previous and the next recorded data.

The daily energy consumption in two years from March 2014 to February 2016 is depicted in fig. 1. The power consumption of campus buildings highly depends on the season in the year. The energy consumption is varied between each time of the year. However, it is repeated in the next year. As aforementioned, the objective of this research is to predict the load consumption with high accuracy and real-time. The ANN is selected for this task.

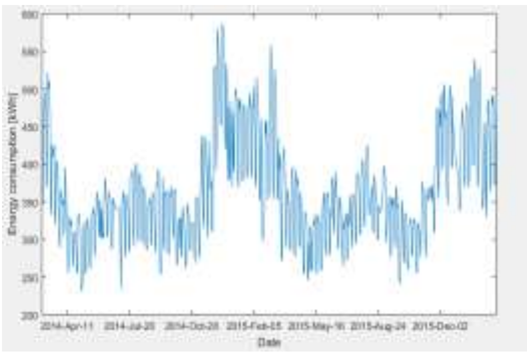

Fig. 1. Daily energy consumption variation for two years

Fig. 2 shows the recorded load profile for each month in two years. In South Korea, the weather is fairly straightforward. The freezing cold, snowy and dry winter is from December to March. The electrical heater is utilized in this time lead to extremely high consumption. In other time in the year, the weather is quite warm. As a result, the energy use is lower. The correlation between outdoor factors will be present in section 2.2.

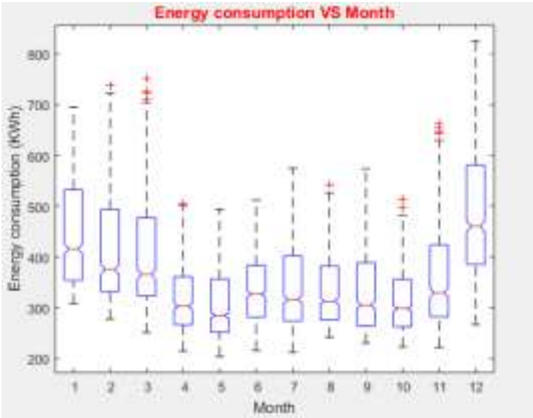

Fig. 2. Distribution of energy consumption for each month

In this study, a short-term load forecasting approach is presented to forecast the electricity consumption of a university campus building. As well know, students and teachers mainly work and study at the campus on weekdays (Monday to Friday) and on working time (9:00 am to 6:00 pm). Consequently, at these time, the load consumption is obviously higher as shown in Fig. 3 and Fig. 4 for a day of the week and each hour of the day, respectively.

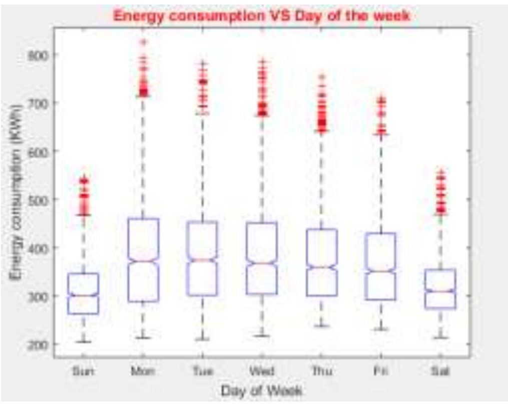

Fig. 3. Distribution of energy use for day of the week

Base on the above analysis, to build a highly accurate forecasting model, the calendar-time should be considered as an important input besides the history load data. Moreover, other outdoor factors also affect to the load, which will be mention in the following section.

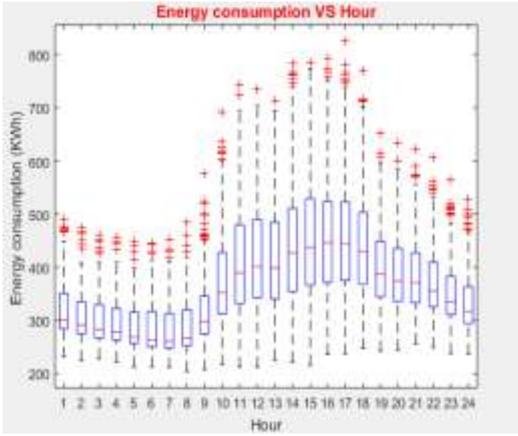

Fig. 4. Distribution of hourly energy consumption

2.2 Other outdoor variables

The weather variables are crucial components of the evolution of building energy systems and minimizing the uncertainty in prediction their evolution can lead to significant saving [24]. Based on this review, the power load is seriously affected by the weather variables such as temperature, humidity, and wind speed. In this paper, these variables are recorded at the local meteorological station. These weather factors are analyzed as follows.

a, Temperature factor First and foremost, temperature has a tremendous impact on power consumption of commercial building on working time because most of the power is consumed by the HVAC system. The temperature curve is represented in fig. 5. This figure shows the major changing of the temperature between each time of the year. The temperature is extremely cold in winter season from December to March. It increases from March and reaches the highest value in summer season (from June to August). The cycle repeats again in the next year.

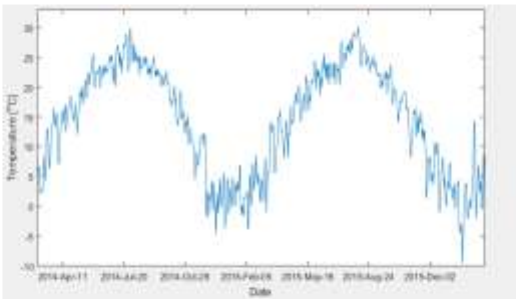

Fig. 5. Daily average temperature curve

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 1078-1091 www.ijaem.net ISSN: 2395-5252

The correlation of daily average ambient temperature with daily average electrical consumption is depicted in fig. 6. The result of correlation shows that the power consumption is highly influenced by the ambient temperature. Due to this reason, the ambient temperature is considered as a significant input for predicting the electric load.

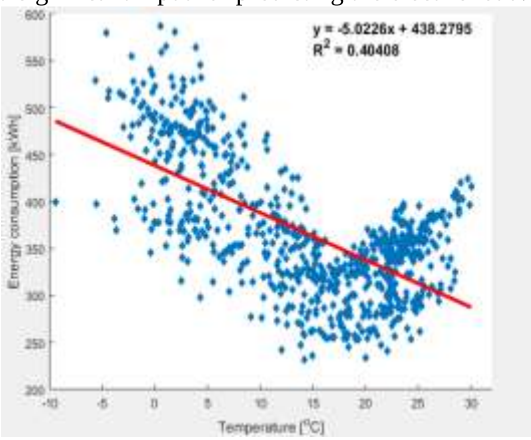

Fig. 6. Correlation between outdoor temperature and energy consumption for a university campus building.

b, Relative humidity factor Relative humidity is measured of the amount of moisture in the air compared to the total amount of moisture that the air can hold. The humidity is influenced by localization and temperature. Fig. 7 shows the humidity curve of the same building. The moisture varies in the large range from 30% to 100%. The warm air in summer season can hold more moisture than the cold air in winter.

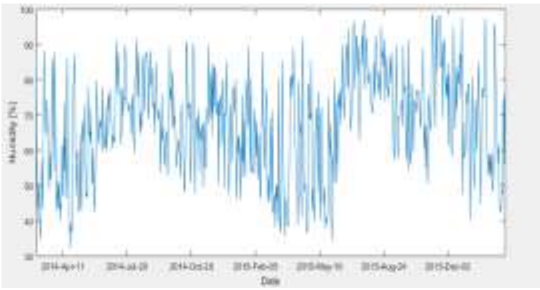

Fig. 7. Daily average humidity curve

The degree of correlation between humidity and energy consumption is presented in Fig. 8, which clearly suggests that the relationship between humidity and energy consumption is not strong. It may be concluded that the humidity factor has little effect on the energy consumption and this factor is not the significant effect in predict model development.

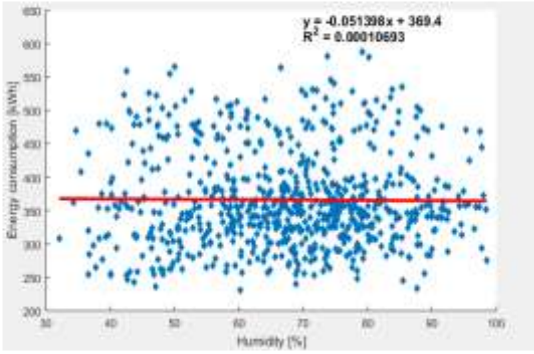

Fig. 8. Correlation of humidity with energy consumption.

c, Wind speed factor Besides temperature and humidity factors, the effect of wind on electricity demand was investigated in [25] using the data from ISONE. They proposed that the wind speed-related variables are able to help improve the temperature-only models and prediction performance. In this section, we examine the actual correlation between wind speed and energy consumption with our practical dataset. The daily average wind speed curve is illustrated in Fig. 9. Fig. 9 shows that the wind speed is highly validation day by day due to the geologically of South Korea, close to the sea. The linear correlation of wind speed and electric power is demonstrated in Fig. 10. The R-square value is extremely low, which indicates that the wind-speed is less effective to the prediction model.

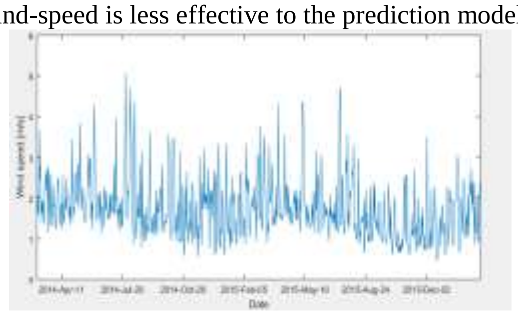

Fig. 9. Daily average wind speed curve

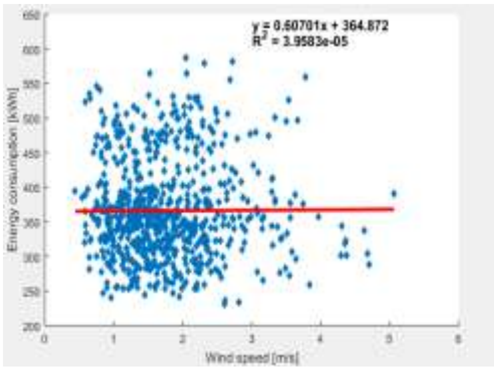

Fig. 10. Correlation of wind speed with energy consumption.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 1078-1091 www.ijaem.net ISSN: 2395-5252

Finally, Table 1 presents the statistic values such as min, max, average, R-squared values of all aforementioned weather variables. Based on this table, the temperature should be mentioned as a critical input while developing the load forecasting model. The temperature will be more investigated in the later section.

Table 1. Statistics of weather variables dataset and relation with energy consumption

	Temperature (^{0}C)	Humidity (%)	Wind speed (m/s)
Min	-11.6000	12	0
Max	35.6000	99	10.1000
Mean	14.4049	67.5173	1.7425
Median	15.6000	69	1.5000
St. Dev.	9.4925	20.0353	1.2320
R^2	0.40408	0.000107	$3.9583*10^{-5}$
N	17544	17544	17544

III. METHODOLOGY

In this section, we present in a detailed framework of implementing process from preprocessing data to output predicted forecasting mode. The sequence of process is depicted in figure Fig. 11.

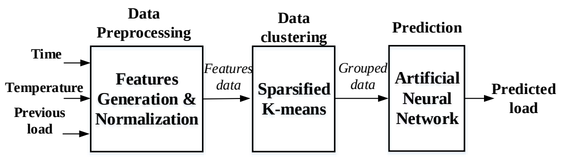

Fig. 11. The proposed method

3.1 Data preprocessing

a, Time features

Based on the analysis in Section 2.1, the electric consumption is critically affected by time index. The specify time variables such as hour of the day (h = 1, 2… 24), day of the week (d = 1 for Sunday, d = 2 for Monday… d = 7 for Saturday), the month of year (m = 1 for January, m = 2 for February… m = 12 for December) are coded by their cosine values, as follows;

$$\hat{h} = \cos\left(\frac{\pi h}{12}\right),\ h = 1,2,\ldots,24. \qquad (1)$$

$$\hat{d} = \cos\left(\frac{\pi d}{3.5}\right),\ d = 1,2,\ldots,7. \qquad (2)$$

$$\hat{m} = \cos\left(\frac{\pi m}{6}\right),\ m = 1,2,\ldots,12. \qquad (3)$$

Furthermore, the daytime, from 9:00 am to 18:00 pm, and nighttime, from 19:00 pm to 8:00 am is denoted as t with the values are 0 and 1, respectively. The days of week are separated to two groups, namely, weekends (Saturday and Sunday) and weekdays (other days) which are represented by w two binary values. From the fig. 2, the load of university campus building is lower on the cold season which is from April to October than the hot season (other months). We denote the cold reason by 1 and hot reason by 0 for s variable. Overall, we have more three binary variables including time of day (t), days of the week (w), the season of the year (s).

b, Temperature features

One of the most impact influenced to load is the temperature. From the previous 24 hour temperatures and current temperature, we propose four variables to represent the temperature consist of current temperature (T(h)), previous hour temperature (T(h-1)), average of temperature in three hours before ($\bar{T}_{3h}$), and the temperature at the same time of previous day (T(h-24)). After computing these four variables, the min-max normalization method is applied to normalize the input data to range [0, 1] as follows.

$$\hat{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \qquad (4)$$

c, Previous load features

The previous load is able to affect the current load [26]. In this paper, we generate five features from the past recorded load which include previous hour load, previous 24 hours load , the average of three closet power loads , the average of six closet power load , and the average load of 24 hours before. These variables are denoted by y (h-1), y(h-24), $\bar{y}_{3h}$, $\bar{y}_{6h}$, $\bar{y}_{24h}$, correspondingly. All of them are calculated and normalized by Eq. 4.

Table 2. Summary of selected features in 24 hours

Features	Selected features	Denoted
Time Features	Normalized hours	$\hat{h}$
	Normalized days	$\hat{d}$
	Normalized months	$\hat{m}$
	Time of days	t
	Weekend and weekday	w
	Season	s
Temperature features	Current temperature	T(h)
	Previous hour temperature	T(h-1)
	Average temperature of three past hours	$\overline{T}_{3h}$
	Temperature at 24 hours before	T(h-24)
Previous load features	Previous load	y(h-1)
	Previous 24 hours load	y(h-24)
	Average of power during three closest past hours	$\overline{y}_{3h}$
	Average of power during six closest past hours	$\overline{y}_{6h}$
	Average of power during 24 closest past hours	$\overline{y}_{24h}$

3.2 Data Clustering Method

The load forecasting problem becomes challenging due to the high diversity of training dataset. Therefore, dividing the training datasets into smaller subgroups based on their common characteristic is considered as a good method to improve the forecasting accuracy. Selecting the features to apply clustering method is crucial steps, which can lead to process in different ways.

The most common method for clustering task with a large amount of data set is K-means clustering approach. The main idea of K-means method is choosing K center points randomly and grouping the input dataset based on the Euclidean distance between the data and the K center point. In [27], W. Wu et al proposed an advanced compression scheme for accelerating k-means clustering namely, sparsified K-means. In this paper, sparsified K-means is selected to group the input data matrix to K group. The process of sparsified K-means approach is as follows with the input data set $X \in \mathbb{R}^{pxn}$.

- Step 1: Initial the number of clustering K and the compression factor γ<1.
- Step 2: The ROS preconditioning transformation is applied to efficiently precondition the data and smooth out large entries in the matrix X before sampling.
- Step 3: Sampling the input data.
- Step 4: Find initial cluster centers via K-means++ [28]
- Step 5: Update the assignments of cluster for each data point and update the cluster center.
- Step 6: The iterative method is applied in Step 2 and 3 until the value of the clustering center no changes. If not, the process continues.

The details of this method is presented in [27]. In our process, the processed features are treated as the input of Sparsified K-means. Each of group will be trained separately using neural network.

3.3 Artificial neural network

Artificial neural network (ANN) has been one of the most effective solutions for prediction energy consumption in the building [20]. The general idea of ANN is inspired by the model of the human brain, which includes dendrite to receive input signal, soma (cell body) to sum all incoming signals to generate input, axon to threshold the sum values and synapses to connection between one neuron and other neurons. A typical ANN comprises different layers consist of input layer, hidden layer and output layer as shown in fig. 12. The activation of a neuron is calculated by summation of the weighted inputs as in Eq. 5.

$$y = f\left(\sum\left(w_{ij}x_j\right)\right) \qquad (5)$$

Where y is the output of the neuron, xj is the input to that neuron, wij is the weight of the connection of the input to the neuron and f is the transfer function. Sigmoid function is selected as the transfer function of ANN due to the ability to handle the nonlinear problems. Sigmoid function is shown in Eq. 6.

$$f(x) = \frac{1}{1+e^{-x}} \qquad (6)$$

The output of the output layer is calculated by Eq. 7. The output is the summation of all output at last hidden layer with the same transfer function.

$$Y = f\left(\sum\left(w_j y_j\right)\right) \qquad (7)$$

The training process of an ANN is the changing weight vector to reduce the error between the real output and the predicted output, which formulated in Eq. 8.

$$E = \frac{1}{2}\sum\left(Y_p - Y_r\right)^2 \qquad (8)$$

Where Yp is the predicted output, Y_r is the real output, and E is the total error.

The most popular training algorithm is the back propagation method which was presented by D. E. Rumelhart et al, in [29].

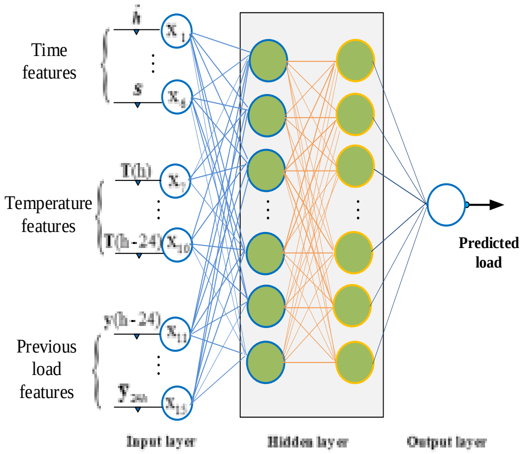

Fig. 12. Inputs-outputs forecasting model.

In comparison with traditional Levenberg-Marquardt back propagation neural networks (LMBP) algorithm, the resilient back propagation (RBP) not only performs faster training and rates of convergence but also has the capability to escape from local minima. Due to these reasons, RBP is adopted for training the predicted model in our study. The main ideal of RBP is to cancel out the harmful effect of the partial derivative on weight-updates. Therefore, only the sign of derivative is considered to indicate the direction where the error function will be changed by the weight-update. An individual update-value $\Delta_{ij}(t)$for each weight, which solely computes the size of the weight-update [30].

$\Delta_{ij}^{(t)}$ is considered for the direction of updating the right.

$$\Delta_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)},\ if\ \dfrac{\partial E^{(t)}}{\partial X_{ij}} > 0 \\ +\Delta_{ij}^{(t)},\ if\ \dfrac{\partial E^{(t)}}{\partial X_{ij}} < 0 \\ 0,\ else \end{cases} \qquad (9)$$

Where the $\frac{\partial E^{(t)}}{\partial X_{ij}}$ express the sum of gradient information for all the patterns and (t) is at the time t.

The neural network with two hidden layers generalizes better performance than those with one [30]. In this paper, we propose to use two hidden layers neural networks with RBP training algorithm. The inputs-outputs architecture employed in this study as shown in Fig.12. As mentioned in this section, each group of data will be trained separately. However, the number of hidden layers is changed from 20 to 30 to investigate the system performance.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

In this section, the proposed methodology is tested on the campus building dataset. The hourly load and temperature dataset used in this test case belong to a database phase of the Smart Energy Campus project, which is now deploying in Chonnam National University, South Korea. The proposed forecasting method is simulated and compared to other methods which based on multiple linear regression, regression tree and SVR methods.

The evaluation metrics most frequently used to assess the performance of a model are the mean absolute percentage error (MAPE) and root mean square error (RMSE) defined by equation Eq. 10 and Eq. 11, respectively.

$$\text{MAPE}=\frac{100}{T}\sum_{t=1}^{T}\left|\frac{y_r(t)-y_f(t)}{y_r(t)}\right| \qquad (10)$$

$$\text{RMSE}=\sqrt{\sum_{t=1}^{N}\frac{\left(y_r(t)-y_f(t)\right)^2}{N}} \qquad (11)$$

Where yr(t) is the real value of load at hour t and yf(t) is the predicted value of load at hour t. T is the total number of hours. N is the number of observed times.

4.1 Comparison with other perdiction methors

We evaluate the prediction performance of different methods including Multiple Linear Regression (MLR), SVM Regression, Regression Tree, ANN without K-means and ANN combined K-means. Notably that all methods use the same feature vector input.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 1078-1091 www.ijaem.net ISSN: 2395-5252

a) One week prediction results

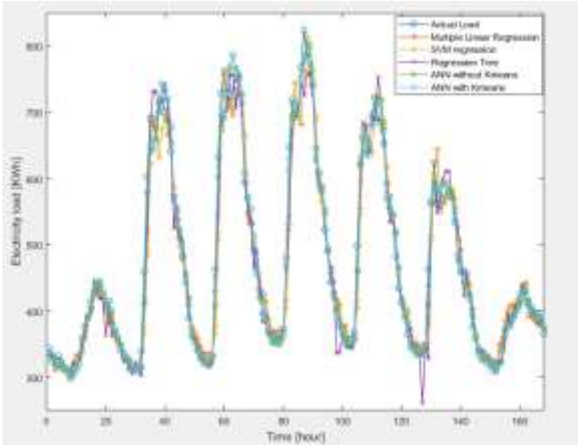

Fig. 13. One week load pattern with different methods

Table A. The performance one week of proposed approach and other methods

Method	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Aver
MLR	2.61	4.61	4.76	4.60	4.48	4.41	3.35	4.12
SVR	2.38	4.69	4.54	4.57	4.32	4.24	2.90	3.95
RTree	2.94	4.49	3.54	3.78	4.61	5.56	2.50	3.92
ANN	2.13	3.39	3.32	2.80	3.01	3.09	1.82	2.79
Proposed method	**2.06**	**3.01**	**2.80**	**2.27**	**2.96**	**2.61**	**1.40**	**2.45**

Table A shows the accuracy of all methods deployed in this paper for each day of the first week from 1/3/2015 to 7/3/2015. The daily and weekly average MAPE using the proposed method are lowest than those of other methods. The best approach reached a MAPE of 1.40%. Taking a closer look at the errors of all day in one week among these methods, we found that the prosed method not only archives the best accuracy in each day but also obtains smallest spreading errors with the ranking from 1.40% to 3.0%. More details in figure Fig. 14 the correlation between real load and predicted load of the proposed method in one week is extremely high with the R-square correlation of approximately 0.99.

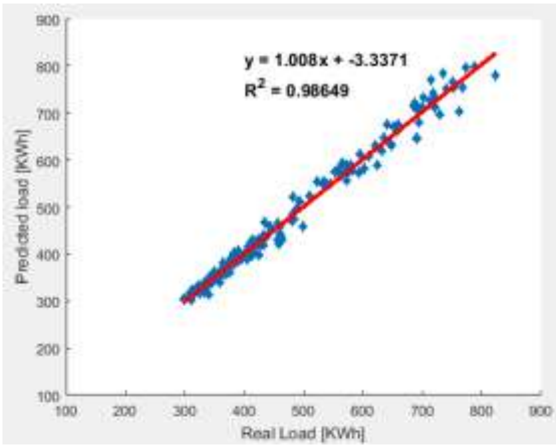

Fig. 14. Correlation between real and predicted value

b) One day prediction results

Based on the prediction time range, the more short-time requires instant and high accurate prediction results. Table B and figure Fig. 15 show the performances and load pattern of one day ahead forecast among methods. We realize that the proposed method not only get the highest accuracy for a whole day but also reach the smallest error at

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 1078-1091 www.ijaem.net ISSN: 2395-5252

every hour.

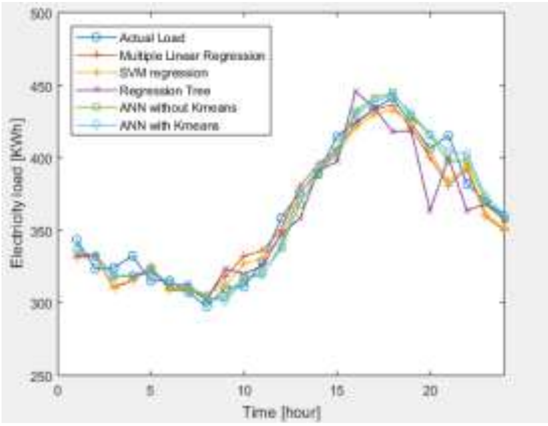

Fig. 15. One day load pattern with different methods

Table B. The performance one day of proposed approach and other method

Method	MAPE	RMSE
RTree_result	2.939	14.417
MLR_result	2.610	11.587
SVR_result	2.384	10.836
ANN_noKmeans	2.132	9.170
Proposed method	**2.061**	**8.861**

4.2 The impact of temperature

As stated in Section 2.2, the forecasting results are mainly affected by temperature feature. We run the evaluation on two different number of input features namely with temperature features and without temperature feature. The prediction results are depicted in Table C for the first week of dataset. As can be seen, the forecasting results are degraded when the system using the input features without temperatures feature. When temperature features are added as the inputs, the performance increases significantly. More specifically, the average error reduces from 3.74% down to 2.45%.

Table C. Impact of temperature to forecasting accuracy

Method	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Average
ANN without Temperature	3.13	4.59	4.34	4.18	3.49	3.22	3.22	3.74
Proposed method	**2.06**	**3.01**	**2.80**	**2.27**	**2.96**	**2.61**	**1.40**	**2.45**

4.3 The impact of different training methods

We also analyze the impact of different training methods. We chose LM training method [29] as the referent due to high accuracy. The forecasting results in one week are presented in Fig. 16. Based on Fig. 16, the proposed method using RBP training algorithm shows better performance. The forecasting result is closer with the actual load in comparison with LM method. More details, the prediction of 24 hours is shown in Fig. 17. We are able to see that from 8:00 am to 10:00 am and 17:00 pm to 21:00 pm, the power load was changed due to office habit in start working time and finish working time. This is also the main difficulty for system to predict with high accuracy at this time. The LM learning method shows more vibration in this time in comparison with RBP method. This result shows that the selected learning algorithm is effected to forecasting performance.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 1078-1091 www.ijaem.net ISSN: 2395-5252

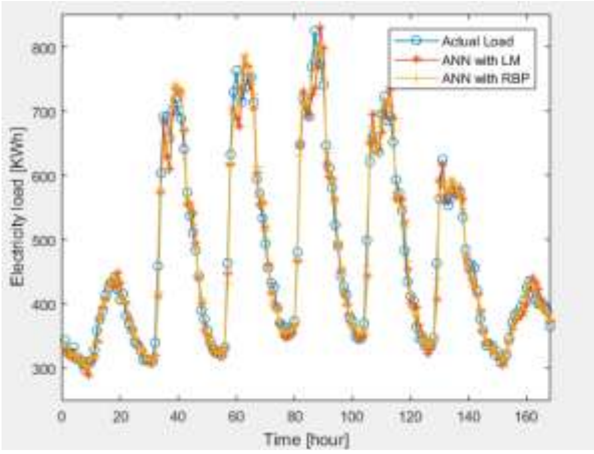

Fig. 16. ANN K-means with different training methods

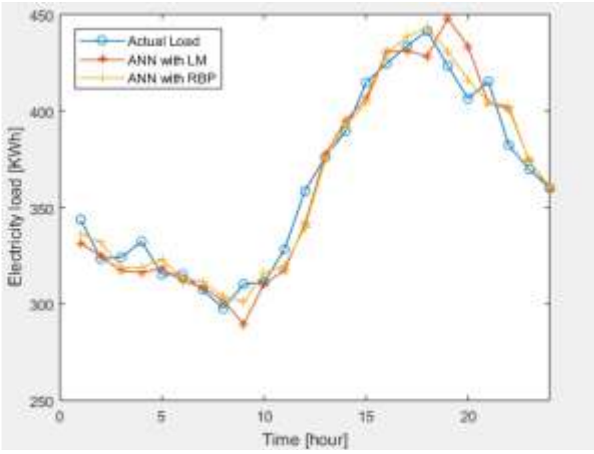

Fig. 17. ANN K-means with different training methods

Table D. MAPE of different training methods

Method	Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7	Average
ANN with LM	2.50	3.55	2.96	3.18	3.87	3.45	2.62	3.16
Proposed method	**2.06**	**3.01**	**2.80**	**2.27**	**2.96**	**2.61**	**1.40**	**2.45**

4.4 Performance of forecasting model for one year

Forecasting model based data-driven always requires re-training model in order to adapt to changing of load profile over the times. This causes much inconvenience in system operating. In this work, we generate the forecasting performance of the proposed model for the next one year and mapping the errors by box-plot of hourly and daily errors as shown in figure Fig. 18 and Fig. 19, respectively. More specifically, 75% of hourly errors has distribution less than MAPE of 5% and 75% of daily errors has distribution less than MAPE of 4%. It shows that by well-combined analysis load characteristic, clustering method and neural network will lead good stable and robust of the forecasting model.

International Journal of Advances in Engineering and Management (IJAEM)
Volume 6, Issue 05 May 2024, pp: 1078-1091 www.ijaem.net ISSN: 2395-5252

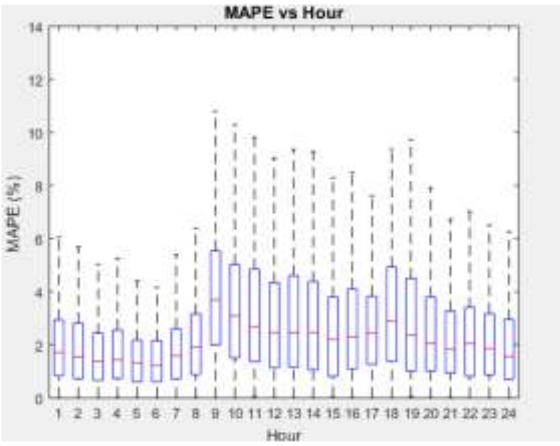

Fig. 18. Distribution of hourly error for next one year

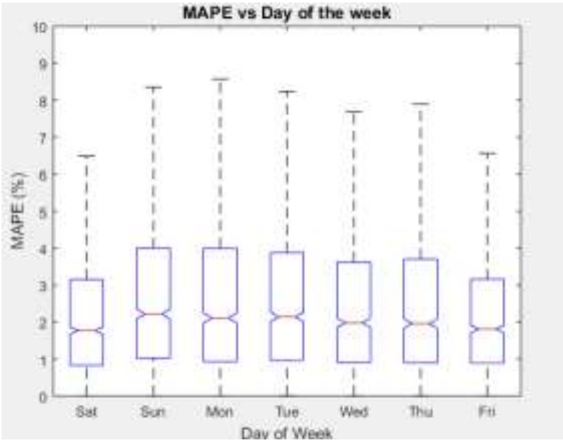

Fig. 19. Distribution of daily error for next one year

V. CONCLUSION

This paper described a detailed process of implementing an artificial neural network combining a clustering method with effective input feature selection in order to predict energy consumption of campus building with highly-nonlinear load profile. The analysis performed showed that neural network combined clustering technique is a powerful method for liberating database phase of microgid energy management system platform due to small-scale observed-data requirement.

The proposed method is computationally simple and also suitable for analyzing a large set of data whose pattern changes over time. The forecasting model is applied to several sets of buildings. From all the experiment results, as the prediction horizon is decreasing, it can be drawn out that ANN based clustering take more superiority in both robustness and accuracy. This enables to apply this method to the real-time energy consumption applications for microgrid.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
ANN	Artificial neural network
ARIMA	Autoregressive integrated moving average
EMS	Energy management system
LMBP	Levenberg-Marquardt backpropagation
MLR	Multiple linear regression
MG	MicroGrid

MAPE	Mean absolute percentage error
ML	Machine learning
MSE	Mean square error
RBP	Resilient back propagation
RMSE	Root Mean Square Error
SVR	Support Vector Regression
SVM	Support Vector Machine
STLF	Short-term load forecasting

REFERENCES

[1]. H. Zhao, and F. Magoulès,“A review on the prediction of building energy consumption,”Renewable and Sustainable Energy Reviews, Vol-16, issue 6, August-2012, pp.3586-3592.

[2]. Aowabin Rahman, Vivek Srikumar, and Amanda D. Smith,“Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks,”Applied Energy, Vol-212, Feb-2018, pp.372-385.

[3]. Korea Energy Economics Institute. Energy Consumption Survey 2014; Ministry of Trade, Industry and Energy: Sejong, Korea, 2014.

[4]. Duk Joon Park, Ki Hyung Yu, Yong Sang Yoon, Kee Han Kim and Sun Sook Kim, “Analysis of a building energy efficiency certification system in Korea,”Sustainability 2015, 7, pp.16086–16107

[5]. L. Hernandez, C. Baladron, JM. Aguiar, L. Calavia, B. Carro, AS. Esguevillas, DJ. Cook, D. Chinarro and J. Gomez, “A Study of the Relationship between Weather Variables and Electric Power Demand inside a Smart Grid/Smart World Framework,” Sensors 2012, 12, pp.11571-11591

[6]. Raza MQ and Khosravi A,“A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings,” Renewable and Sustainable Energy Reviews 2015, 50, pp.1352–1372.

[7]. Liu K,“Comparison of very short-term load forecasting techniques,” IEEE Transaction on Power Systems 1996, 11, pp877–882.

[8]. Huang SR, “Short-term load forecasting using threshold autoregressive models,” IEEE Proceedings-Generation, Transmission and Distribution, 1997, 144(5), pp.477–481.

[9]. J. Nowicka-Zagrajek and Rafal Weron, “Modeling electricity loads in California: ARMA model with hyperbolic noise,” Signal Processing 2002, vol. 82, pp.1903-1915.

[10]. JF. Chen, WM. Wang, CM. Huang, “Analysis of an adaptive time-series autoregressive moving average (ARMA) model for short-term load forecasting,”. Electric Power Systems Research, 1995, 3, pp.187-196

[11]. Juberias G,“A new ARIMA model for hourly load forecasting,” Proceedings of the 1999 IEEE transmission and distribution conference.

[12]. V. Bianco, O. Manca and S. Nardini, “Linear regression models to forecast electricity consumption in Italy,” Energy Sources, 2013, 8, pp.86-93

[13]. Tao Hong, Pu Wang and H. Lee Willis, “A naive multiple linear regression benchmark for short term load forecasting,” Proceedings of the 2011 IEEE power and energy society general meeting.

[14]. Dudek G, “Multivariate Regression Tree for Pattern-Based Forecasting Time Series with Multiple Seasonal Cycles,” ISAT 2017, Vol 655, Springer.

[15]. RK. Jain, KM. Smith, PJ. Culligan and JE. 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 2014, Vol 123, pp.168-178

[16]. MA. Mat Daut, MY. Hassan, H. Abdullah, HA. Rahman, MP. Abdullah and F. Hussin, “Building electrical energy consumption forecasting analysis using conventional and artificial intelligence methods: A review,” Renewable and Sustainable Energy Reviews 2017, 70, pp.1108-1118

[17]. Wei Yu, B. Li, Y. Lei and M. Liu,“Analysis of a residential building energy consumption demand model,”

Energies 2011, 4(3), pp.475-487.

[18]. RB. Carlos, EE. Guillermo, AB. Carlos, RP. Carlos, RG. Javier,“Upgrade of an artificial neural network prediction method for electrical consumption forecasting using an hourly temperature curve model,” Energy and Buildings 2013, 60, pp38–46.

[19]. Mai W, “Electric load forecasting for large office building based on radial basis function neural network,” Proceedings of the 2014 IEEE PES general meeting conference & exposition

[20]. Chirag Deb, LS. Eang, J. Yang and M. Santamouris, “Forecasting diurnal cooling energy load for institutional buildings using Artificial Neural Networks,” Energy and Buildings 2016, 121, pp.284–297.

[21]. Ervin Ceperic, Vladimir Ceperic and Adrijan Baric, “A Strategy for Short-Term Load Forecasting by Support Vector Regression Machines,” IEEE Transactions on Power Systems 2013, 28, pp4356-4364

[22]. GP. Zhang, “Time series forecasting using a hybrid ARIMA and neural network model,” Neurocomputing 2003, 50, pp.159–175.

[23]. Hongzhan Nie, Guohui Liu, Xiaoman Liu and Yong Wang, “Hybrid of ARIMA and SVMs for Short-term Load Forecasting,” Energy Procedia 2012, 16, pp.1455-1460

[24]. D. Lazos, A.B. Sproul, M. Kay, “Optimisation of energy management incommercial buildings with weather forecasting inputs: a review,” Renewable and Sustainable Energy Reviews 2014, 39, pp.587–603

[25]. Jingrui Xie and Tao Hong, “Wind Speed for Load Forecasting Models,” Sustainability 2017, 9, pp.795-806

[26]. L. Hernandez, C. Balandron, JM. Aguiar, B. Carro, AJ. Sanchez-Esguevillas and J. Lloret,“Short-term Load Forecasting for Microgrids Based on Articifial Neural Networks,”Energies 2013, 6, pp.1385-1408.

[27]. PA. Fahad and Stephen Becker, “Preconditioned Data Sparsification for Big Data with Applications to PCA and K-means,” IEEE Transactions on Information Theory 2017, Vol. 63, pp. 2954-2974

[28]. David Arthur and Sergei Vassilvitskii, “k-means++: The Advantages of Careful Seeding,” Proceeding of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms, 2007

[29]. D.E. Rumelhart, G.E. Hinton and R.J. Williams, “Learning representations by back-propagating errors,” Nature 323 (1986), pp533–536

[30]. M. Riedmiller and H. Braun,“A direct adaptive method for faster backpropagation learning: The RPROP algorithm,” Proceedings of IEEE International Conference on Neural Networks, San Francisco, CA , USA, 28 March–1 April 1993, pp.586-591.