International Journal of Marketing and Technology
Vol. 12 Issue 12, December 2022,
ISSN: 2249-1058 Impact Factor: 6.559
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

METHODS AND FEATURES FOR SHORT-TERM LOAD PREDICTION IN A COMMERCIAL BUILDING

Abhinav Gupta
B.Tech. Computer Engineering
Trinity College Dublin

Abstract

The electricity system is undergoing change. There will be improvements in the grid's efficiency, resilience, and safety thanks to smart grid technology and demand response systems. Accurate load forecasting is a crucial part of power system management. When a utility company has accurate short-term load forecasting, it may better manage its resources and implement control measures to maintain a stable energy supply and meet customer demand. The purpose of this work is to develop a system for estimating a commercial building's future electrical consumption. The second objective is to determine which types of information, such as external and internal temperatures, dates, and building occupancy, are most useful for load forecasting in buildings. Simple experiments using MLR, MLP, and SVR models demonstrate the suggested prediction method's high accuracy and minimal computational cost.The findings of the case study demonstrate the excellent accuracy and cheap computational cost of the suggested forecast method.

Keywords: Load prediction;MLP Regression;SVR and Multiple linear regression

1. Introduction

Because of its vital role in global trade and production, energy must be made more affordable and environmentally friendly. Transport losses, centralization of power generation, reliance on huge power plants, etc. are only a few of the problems with the traditional electric grid. The goal of the new electrical power grid, dubbed the "smart grid," is to utilize decentralized generation and load management to shape the load curve. Smart grids will make the network sturdy, dependable, efficient, and adaptable. Adjusting power generation to a user's usage in real time is one of the most pressing difficulties facing utilities in light of these novel ideas. If you underestimate demand, you'll have to raise prices to meet it, but if you overestimate it, you'll squander resources. This modification will now occur in more compact settings, such as micro grids, as opposed to the larger ones that were typical in the past. Currently, the most effective method for balancing the grid's two sides is based on predictions of electric load. Seeing as the construction industry is responsible for about 40% of all CO2 emissions **(Casals, 2006),** STLF in this sector is crucial for cutting down on energy usage. When it comes to electric power, city load curves are distinct from building ones. More noise, non-linearity, and variation can be seen in building load curves. The more data points there are, the harder it is to aggregate them into meaningful patterns. On top of that, there are numerous building types, including residential and commercial structures. Consumption in non-residential buildings, such as those used for university administration, exhibits daily, weekly, and seasonal patterns. Over the course of the week's nonworking hours, especially on the weekends and holidays, consumption drops dramatically. Daily patterns emerge from the user's habits and routines. On a weekly scale, there are consistent consuming habits.

Seasonal changes follow predictable cycles throughout the year. Additionally, there are liminal periods of consumption during the morning and evening hours of the workday, as well as during the lunch and breakfast hours. Most buildings older than 10 years use manual, arbitrary criteria to manage their HVAC systems. The calendar and building occupancy rates can also play a role in estimating energy use. Knowing what data is most critical to measure and collect is another crucial challenge in electricity forecasting. Although future load forecasting is vital for smart building management, this study distils the process down to its barest essentials in order to minimise any inaccuracies in the prediction of consumption. Difficulties in parameterization, variable selection, and over-fitting are no longer an issue because to the development of modern computing technology. The first large group contains the work of **(Lam et al. 2010),** in which principal component analysis is used to derive a new climate indicator called Z. The MLR is then applied to this index to predict energy use. Concerning ANN, the study by **(Karatasou et al 2006)** suggests combining consumption and meteorological data to foretell load consumption in workplaces. It is concluded from this research that while wind speed and humidity are not crucial, temperature and sun radiation are. In a study using fabricated consumption and climate data, ANN's simplicity was cited as its key strength. Using the same data format as the previous two examples, but this time for a hotel, **(Li, and Su, 2010)** propose via an ANN named ANFI, optimised using GA. The efficiency of ANNs is boosted as a result of this. A modest number of days with similar work activity and temperature are chosen to train an artificial neural network (ANN) for the instance **(Ortega et al., 2011),** campus energy consumption is predicted based on expected values of temperature.

2 Objective of the study

The goal is to achieve maximum accuracy while using as few data attributes and instances as possible. This report compiles information on the 10-year-old office building's electricity use, occupancy rates, internal climate, and more. Then, the performance of three distinct models was compared to determine a clear victor. The three selected models—MLR, MLP, and SVR—are all industry standards with satisfactory performance. This experiment has been conducted with actual data obtained from a WSN.

3 Methodology

3.1. Models

We investigate three widely used models (MLR, MLP, and SVR) to determine which works best.

3.1.1. Multiple linear regression

In the MLR case, the system is modelled as a linear equation with numerous independent variables and a single dependent variable **(Williams, 2013).** The dependent variable at a given time period can be explained by a set of independent factors. Model description is given in the form of an equation, as seen in Equation (1)

$$y(t) = \beta_0 x_0(t) + \beta_1 x_1(t) + \cdots + \beta_n x_n(t) + \beta(t) \ldots \ldots \ldots . . (1)$$

3.1.2. Multilayer perceptron

The multilayer perceptron (MLP) **(Bui et al., 2018)** is a stack of densely connected nodes. There are three distinct layers that make up an MLP: the input layer, the hidden layer, and the output layer.

Adjusting the weights of each node is crucial to the learning process. In terms of popularity, the EBP can be attributed to its status as the most widely used learning method.

3.1.3. Support vector regression

SVM aims to locate a hyper-plane that may be used to classify data **(de Mello, 2018).** In order to discriminate between two classes that are not linearly separable, data are transformed using kernel functions and relocated to a high dimensional feature space where data are linearly separable. Eq. (2) specifies the SVR function in full:

$$f(x) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i) k(x_i, x) + b \ldots \ldots . (2)$$

Each of the three levels in the commercialstructure has a capacity of 4000 m^3. Construction began in 2003. The front face is a whopping 1836 square metres in size, with 630 of that being glass. The HVAC system relies on a gas boiler heating system with 824 kW of capacity. A combination of fancoils and a compression refrigeration system manufactured by Climaveneta make up the 618 kW cooling system. The thermostat requires manual tuning. There is a 20°C setting in the winter and a 26°C setting in the summer. The HVAC system monitors indoor temperatures in offices and classrooms, and adjusts fan speeds accordingly to maintain a constant, predetermined temperature.

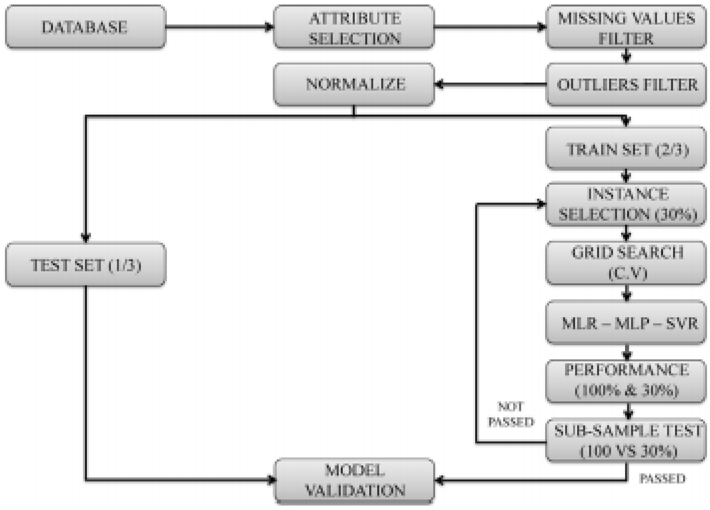

Fig. 1. Block diagram of the process(Casals, 2006)

Figure 1 provides a summary of the suggested model. To begin, we perform a feature selection by discarding unnecessary characteristics. In the following section, we remove anomalies and missing data. The next step is to normalise the data (66% of the total) and then split it into test data (33%). By use of a random sampling procedure, we reduce the size of the training data set to 30%. With this 30%, we run a GS to determine the models' parameters and effectiveness. We then carry out a validation on a smaller subset of the data.

It used random sub-samples instead of training the model on the full dataset due to the significant computational cost of doing so and conduct a simple test with the results of each subsample to see if they still follow the same pattern of dispersion as the original population. then utilise the parameters that yield these three best outcomes to compute the models' and the dataset's performances. Afterward, make two separate rankings, each with three spots, based on how well each dataset performed. After this is complete, we verify that the sample's ranking for each parameter set matches the baseline dataset's ranking. If this is the case, then the sample is representative, and these values should be used for the parameters. If not, the procedure will repeat with a new subsample.

4. Experimental results

The studies have been conducted on a "PC with an Intel Core i7-4500U processor and 8 GB of DDR3 RAM" using Weka programme**(Kaunang, & Rotikan, 2018).** Indicators used to evaluate performance, together with the scenarios studied, and the final results, will be discussed below.

4.1. Error indicator

Several methods exist to compute the model's quality, the most well-known of which are the (MAPE) and the (CC) in the existing study.

For comparing models, utilise the MAPE which is independent of the size of the units used The CC is a linear correlation coefficient that compares actual and expected output.

4.2. Results

Here compare and contrast the predictions made by the MLR, MLP, and SVR simulations under a various conditions. We give four factors for each scenario: model performance (MAPE and CC), to calculate time, and features. The selected model is the one that provides the optimal compromise between various metrics. In addition to the predictability, computational time, and attribute usage, the clarity of the model is an intriguing challenge. The MLR is easily understandable, while the MLP and SVR are not.

The first experiment's results showed poor performance indicators and lengthy computation times due to noise and redundant data. Although an SVR model can produce accurate forecasts, the extra work required to process the raw data is noticeable.

Model	CC	MAPE	Computation time
MLR	0.1755	24.3%	3 s
MLP	0.2463	23.72%	1843 s
SVR	0.964	14.32%	7546 s

Table 5 attributes and instances of all 3 models

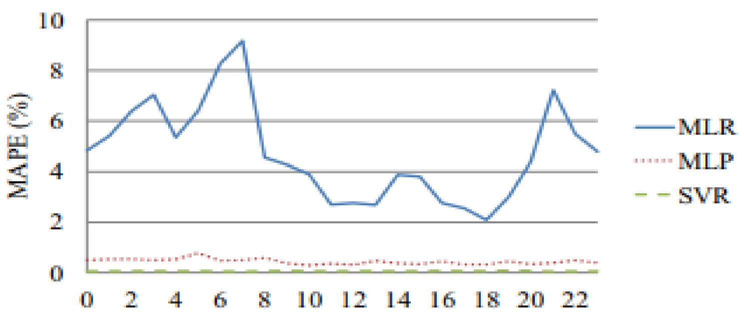

Figure 6 is a MAPE v/s. Time chart for the three models

Due to the unpredictability of security and cleaning services by the hour, we discover the worst accuracy in nighttime forecasting in the MLR instance, and to a lesser extent in the MLP and SVR examples. There is the least reliability in making predictions at the beginning and conclusion of the workday, as well as during lunch.

5. Discussion and conclusions

We compared three models for distinct groups of characteristics while keeping in mind a wide range of parameters. Using characteristics like temperature, date, and occupancy, MLR and MLP models perform at their highest levels of efficiency.

Predicting electricity generation and consumption in the building sector is essential in utilities' efforts to strike a balance between the two. To that end, we have set up a WSN to gather information on the building's infrastructure in order to develop an STLF model for a commercial office building that is 10 years old. The most important attributes were identified by an examination of meteorological, indoor ambient, calendar, and actual building occupancy data. Next, we put the MLR, MLP, and SVR models through their paces using a variety of data points. To conclude, the SVR model, which uses just temperature and actual occupancy data, provides the optimal trade-off between precision and computing effort. The cost of setting up a WSN, the size of the database, and the amount of processing required can all be minimised if unnecessary attributes are removed. The data mining method will hence necessitate less robust infrastructure.

Reference

X.G. Casals, (2006) Analysis of building energy regulation and certification in Europe: their role, limitations and differences, Energy Build. 38 (5) 381–392

S. Karatasou, M. Santamouris, V. Geros, (2006) Modeling and predicting building’s energy use with artificial neural networks: methods and results, Energy Build. 38 (8) 949–958

J.C. Lam, K.K.W. Wan, S.L. Wong, T.N.T. Lam, (2010) Principal component analysis and long-term building energy simulation correlation, Energy Convers. Manage. 51 (1) 135–139.

K. Li, H. Su, (2010) Forecasting building energy consumption with hybrid genetic algorithm-hierarchical adaptive network-based fuzzy inference system, Energy Build. 42 (11) 2070–2076

G. Escrivá-Escrivá, C. Álvarez Bel, C. Roldán-Blay, M. Alcázar-Ortega(2011), New artificial neural network prediction method for electrical consumption forecasting based on building end-uses, Energy Build. 43 (11) (2011) 3112–3119

M. Hall, E. Frank, G. Holmes, (2009) The WEKA data mining software: an update, ACM SIGKDD Explore. Newslett. 11 (1) 10–18

de Mello, R. F., & Ponti, M. A. (2018). Statistical Learning Theory. *Rodrigo Fernandes de Mello*, 75.

Kaunang, F. J., & Rotikan, R. (2018, October). Students' academic performance prediction using data mining. In *2018 Third International Conference on Informatics and Computing (ICIC)* (pp. 1-5). IEEE.

Williams, M. N., Grajales, C. A. G., &Kurkiewicz, D. (2013). Assumptions of multiple regression: Correcting two misconceptions. *Practical Assessment, Research, and Evaluation*, *18*(1), 11.

Bui, D. T., Nhu, V. H., & Hoang, N. D. (2018). Prediction of soil compression coefficient for urban housing project using novel integration machine learning approach of swarm intelligence and multi-layer perceptron neural network. *Advanced Engineering Informatics*, *38*, 593-604.