**Prediction of building energy consumption**

[ Energy AI 김소희 ]

1. **Prediction models for forecasting building energy consumption** [1]

Prediction of building energy consumption is crucial for improved decision making towards reducing energy consumption and CO2 emissions, because it can assist in evaluating different building design alternatives and building operation strategies (in terms of their energy efficiency) and improving demand and supply management. However, building energy consumption prediction remains to be a challenging task due to the variety of factors that affect the consumption such as the physical properties of the building, the installed equipment, the outdoor weather conditions, and the energy-use behavior of the building occupants.

There are several popular methods used for forecasting building energy consumption. It can be categorized into two main approaches, engineering (physical) modelling approach and data-driven approach. Physical models (also known as white-box models) rely on thermodynamic rules for detailed energy modelling and analysis. Examples of building energy simulation software that utilize physical models include EnergyPlus, eQuest, and Ecotect. These types of software calculate building energy consumption based on detailed building and environmental parameters such as building construction details; operation schedules; HVAC design information; and climate, sky, and solar/shading information [4]. However, some of such detailed data may not be available to the users at the time of simulation. Failure to provide accurate input can result in poor prediction performance

On the other hand, data-driven building energy consumption prediction modelling does not perform such energy analysis or require such detailed data about the simulated building, and instead learns from historical/available data for prediction. They correlate the energy consumption or energy index with the influencing variables. Data-driven prediction has gained a lot of research attention in recent years, despite its possible limitations. Developing a data-driven model, typically, consists of four primary steps: data collection, data preprocessing, model training, and model testing. SVM, ANN, decision trees, and other statistical algorithms are the most commonly-used supervised machine learning algorithms for model training.

Despite the importance of data-driven approaches, data-driven energy consumption prediction has two main limitations. First, data-driven prediction models may not perform well outside of their training range. For example, a model that was trained by learning from a limited dataset (e.g., data collected from a small set of buildings) may not perform well outside of the training data (e.g., different types of buildings in terms of physical properties, operation strategies, weather conditions, occupant behavior, etc.). Second, data-driven prediction models are black-box models – their internals are not known. A black-box model may provide sufficient prediction accuracy but may be limited in providing a detailed understanding of the different parameters and its behavior in terms of energy consumption.

Hybrid or grey-box modelling approaches, on the other hand, offer a combination of physical and data-driven prediction models, thereby leveraging the advantages and minimizing the disadvantages of both approaches. In grey-box models, some internal parameter and equations are physically interpretable. Grey-box models may also show better performance compared to black-box and white-box models.

So, I will discuss about 2 paper which is each related to data-driven model (black-box model) [2] and a hybrid model (grey-box model) [3]. Chae et al. [2] proposes a short-term building energy usage forecasting model based on an Artificial Neural Network (ANN) model with Bayesian regularization algorithm and investigates how the network design parameters such as time delay, number of hidden neurons, and training data effect on the model capability and generality. Dong et al. [3] developed a hybrid model, which couples a data-driven model and a thermal network model, for predicting the total and non-AC energy consumptions of residential buildings and compared its prediction performance to ANN-, SVM-, LSSVM-, Gaussian mixture model (GMM), Gaussian process regression (GPR)-based models.

1. **Artificial neural network model for forecasting sub-hourly electricity usage in Commercial buildings** [2]
   1. **Introduction**

Forecasting electricity load, especially for commercial and industrial buildings has become one of the important topics recently, to be able to better manage energy usage. Although the forecast time horizon can range from minutes to years, the short-term load forecast (STLF), especially for a period shorter than a day, has been more of an interest in the perspective of buildings because the utility prices may change by seasonality, time-of-use in on/off peak period, and contract demand.

Previous studies on STLF for the sub-hourly electricity consumption of buildings are limited. Escriva et al. [4] proposed STLF model using ANNs, but this model requires an entire whole year’s dataset and the performance may not be stable when the energy consumption pattern has large daily or annual variability. Therefore, it is useful to explore a model that can perform well under more general setting, in particular, not requiring a large amount of input data for forecasting electricity usage of buildings. For this need, they developed a short-term load forecasting model using data mining and machine learning technique while assuming limited availability of data. In particular, they investigate ANNs model to predict the energy consumption of a commercial building complex.

* 1. **Methodological approach**

 All data set for this study was obtained from a building management system (BMS) of a commercial office building complex, and the data are periodically pulled into a relational database (IBM DB2TM). The site consists of three office buildings in urban area, each of which has different number of floors; five in building 1 (BLDG1), four in building 2 (BLDG2), and two in building 3 (BLDG3). A total floor area of 15,224 m2 spreads over typical office area, small laboratories, cafeteria, parking garage, and small gymnasium (Fig. 1). Although the buildings are separated, they are all managed by one utility billing system.

Figure 1 Typical floor plans of the case study building complex

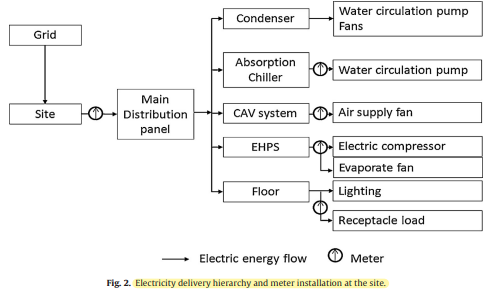
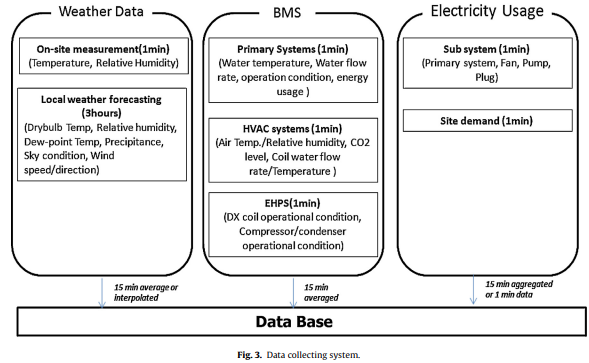
Two absorption chiller systems provide chilled water in summer and hot water in winter for a constant air volume (CAV) system for each floor of BLDG1 and fan coil units (FCU) of perimeter zones of each floor of BLDG2. All three buildings have electric heat pump (EHP) systems with multi-indoor units. EHP systems are supplementary system incorporated with CAV systems for BLDG1 but operate as the main air-conditioning system for BLDG2 and BLDG3. BMS system monitors operational conditions of both primary/secondary system and EHPs. The system also controls all secondary system operation, whilst EHPs are locally controlled by occupants’ indoor thermal demand. For the electricity usage monitoring, one main electric meter and several sub-meters are installed as illustrated in Fig. 2. The main meter measures electricity usage, both the instantaneous power in kW with a minute interval and aggregated electricity usage at every 15 minutes in kWh.

Figure 2 Electricity delivery hierarchy and meter installation at the site

The relational database management system (RDBMS) has been used to collect and store environmental variables, BMS data, and electricity usage as illustrated in Figure 3. Although the data set has over 1000 data points with 20 different measurement types, some variables may not eventually be used for electricity usage forecast. Thus after taking the data availability into account, the predictors are divided into three categories: environmental data, time indicator, and operational condition, as illustrated in Table 2.

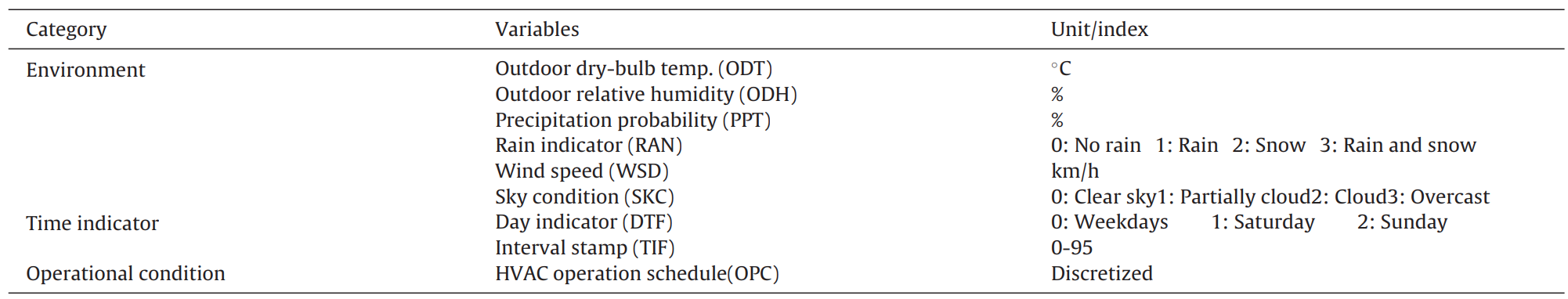


Figure 3 Data collecting system

Table 1 Potential predictor variables

The data include outdoor air temperature, relative humidity, wind speed and direction, sky condition, and precipitation(강수량) type in every 3-hour interval for the next 72 hours. The forecast data is updated eight times a day.

The potential predictors are nine independent variables. However, those variables influence the electricity usage of the building in a different way. If some variables in the input data set are irrelevant to the output, it decreases model accuracy, stability, and effectiveness. Therefore, it is necessary to pre-screen the variables by identifying the important variables from the input data set.

Random forests algorithm [5] was used to assess the importance of variables by measuring the candidate parameters in terms of their impacts on the response of prediction. The permutation importance and node impurity were used to select variables in this study. According to permutation importance and node impurity, the operational condition was one of the most important factors. It shows that the operational condition of the secondary system is useful to capture the actual activity in the building such as occupancy and electricity consumptions of lighting systems and receptacles.

Five variables are ranked highly in their importance; the operational condition, time indicator, day type, outdoor dry-bulb temperature, and outdoor relative humidity. These five variables are selected as input attributes, together with previous electricity usages.

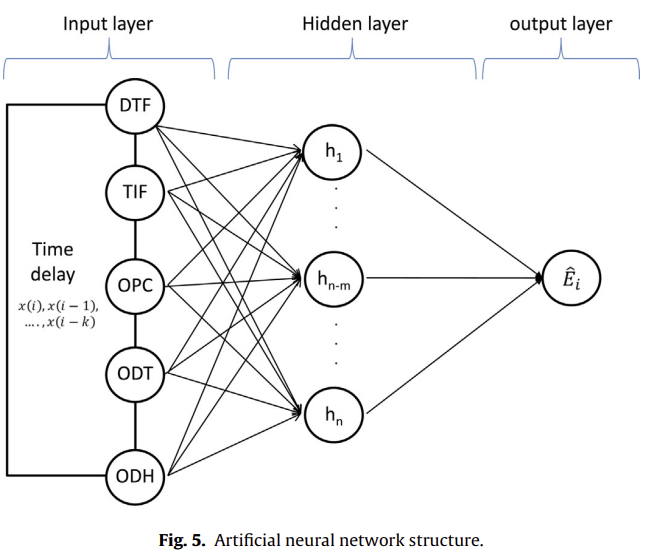
Although many machine-learning algorithms are available, the choice of the specific method to use is not trivial and depends strongly on the specific application and data availability and type. They examined nine different machine-learning algorithms and chose the one that performed best, the ANN. The ANN model in this study has a conventional multi-layered feedforward network using a backpropagation algorithm as illustrated in Figure 5. A Bayesian regularized neural network model with Levenberg–Marquart(LM) backpropagation algorithm is employed for the training process to improve the generalization of model. In this approach, the objective function includes both the conventional error function and the weight decay components or penalty term. The weights and biases in the model are assumed to be random variables with Gaussian distribution and the regularization parameters in the objective function can be optimized by using Bayesian rules.

Figure 4 Artificial neural network structure

A test and validation procedure of the ANN model was conducted under several predictor conditions and data implementations with time delays during the training period. The 15-minute interval data set and highly ranked five predictor variables were collected from July 1st to July 31st, 2012. By removing data from one day, which has sensor and meter malfunction, 30 days, of which 22 are weekdays and 8 are weekend days, were used. Total of 2880 data points for six input variables, including the electric usage, were used to train the model. The network model selected the training input data set in a random, while three weekdays with a new data set (August 1–3, 2012), due to the weather forecasting time scope, were used to evaluate the out-of-sample testing.

* 1. **Results and discussions**

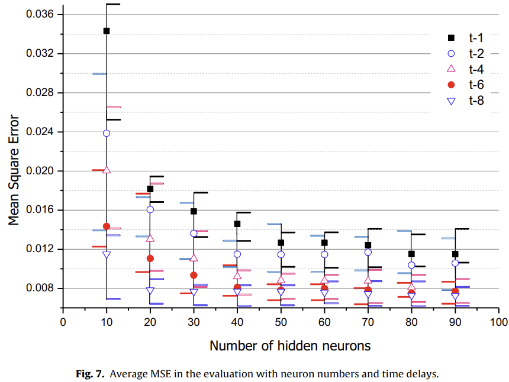
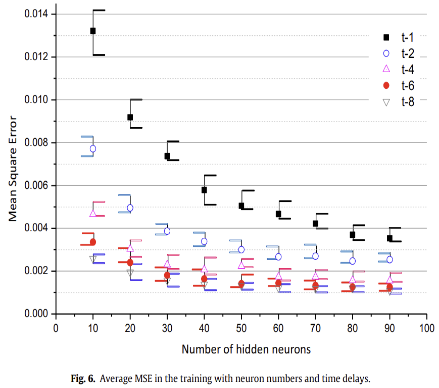


Figure 5 Average MSE in the training(left) and evaluation(right) with neuron numbers and time delays.

Figure 5 shows average and min/max error range of MSE between the actual energy consumption and the model, presented values in the training and evaluation procedures. The total neuron number in the hidden layer is varied from 10 to 90. The input-feedback time delay (i-k) is also parameterized, for the time step k = 1, 2, 4, 6, 8 (up to 2 hours). Although the training algorithm has the regularization function, the network requires a relatively large number of neurons and time delays to provide a stable performance. Considering the model complexity and computation time, it is reasonable to have 50 neurons in the hidden layer and have t-6 time delay for input variables and feedback for the network model in this study.

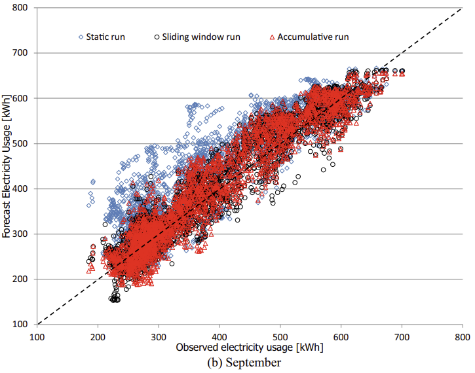
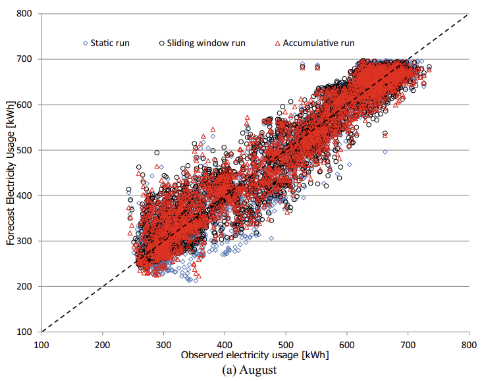


Figure 6 Comparisons of actual observed and forecasted electricity usage.

To verify the forecasting performance of the developed model, two months of data, August and September in year 2012, were used. Three training methods were considered: static, accumulative, and sliding windows. For the static training, the model is trained using four weeks of data in July, and it forecasts for August and September without retraining the model with newly available data. Accumulative training, an adaptive training method, uses accumulated data set from the first day of July to the day before the target day and retrained on a daily basis. The sliding windows method uses a fixed training data window size (four weeks) and the window is shifted by a day by removing the first day of the old training set and adding the new measurements into the data set. The networks are also retrained daily with the new training data

As illustrated in Figure 6, the overall prediction performance of each training type is similar to each other in August. The coefficient of determination (R-square) values is 0.904, 0.912, and 0.902 for each method. The result of September shows that accumulative and sliding windows are similar to August, but static training was not good as in August(R-square = 0.798).

테이블이(가) 표시된 사진

자동 생성된 설명

Table 6 presents the statistical summary of CV(RMSE) and its standard deviation() on daily level. All three run types of training set have a similar performance in August

1. **A hybrid model approach for forecasting future residential electricity consumption** [3]
   1. **Introduction**

**References**

[1] Amasyali, Kadir, and Nora M. El-Gohary. "A review of data-driven building energy consumption prediction studies." *Renewable and Sustainable Energy Reviews* 81 (2018): 1192-1205.

[2] Chae, Young Tae, et al. "Artificial neural network model for forecasting sub-hourly electricity usage in commercial buildings." *Energy and Buildings* 111 (2016): 184-194.

[3] Dong, Bing, et al. "A hybrid model approach for forecasting future residential electricity consumption." *Energy and Buildings* 117 (2016): 341-351.

[4] Escrivá-Escrivá, Guillermo, et al. "New artificial neural network prediction method for electrical consumption forecasting based on building end-uses." *Energy and Buildings* 43.11 (2011): 3112-3119.

[5] Breiman, Leo. "Random forests." *Machine learning* 45.1 (2001): 5-32.