Homogeneous Ensemble Model for Building Energy Prediction: A Case Study Using Ensemble Regression Tree

Zeyu Wang and Ravi Srinivasan, University of Florida
Yueren Wang, Microsoft Corporation

ABSTRACT

The building sector accounts for the largest proportion of U.S. primary energy use. Therefore, improving building energy efficiency is of great importance in achieving energy conservation and overall sustainability. Building energy prediction is an efficient tool for building energy management, system fault detection, and demand response control. Efforts toward building energy prediction are meaningful and crucial for building energy efficiency and building sustainability. This paper discussed the development, test, and validation of a homogeneous ensemble model for short term building energy use prediction. A homogeneous ensemble model works as a prediction enhancement approach that can reduce the variance and the bias of learning algorithms. In this study, an ensemble bagging tree (EBT) was proposed as an ensemble model for building energy prediction. A regression tree (RT) was picked as a non-ensemble model to map the nonlinear relationship between input and output data and bagging, used as a data partition strategy, was applied to subsample training data into multiple subsets. In this article, the input data was comprised of weather conditions, day types, and time of day, while the output data was hourly building level electricity usage. The ensemble model was trained and tested with data collected from an institutional building. A comparison between the proposed EBT model and the conventional RT was conducted after EBT development to investigate the difference in prediction accuracy and instability between these two methods.

Introduction

There is worldwide increase of energy demand in the past few decades because of the growth of population as well as the rise of living standard (U.S. Energy Information Adminstration 2016). Meanwhile, because of the worldwide natural resource depletion, the energy supply has run into a bottleneck that has resulted in global energy crisis. In addition, the excess use of energy has created severe environmental issues such as climate change, pollution, and resource depletion (Edenhofer 2015). Therefore, energy conservation is of great importance for relieving global energy crisis and achieving sustainability.

As a major energy consumer, the building sector accounts for 41% of U.S. primary energy usage in 2014, which is 46 and 35 percent more than transportation sector and industrial sector, respectively (Energy Information Administration 2015). Because of the great amount of energy that buildings consumed, improving their energy efficiency is essential for energy conservation. During the past few decades, building energy prediction has attracted many researchers' attention because of the important role it plays in improving building energy efficiency. Accurate building energy prediction is currently utilized for several purposes: optimizing energy control systems or strategies for a building, determining cost-effective retrofit for existing buildings, detecting and diagnosing faults of building system, shifting building energy to off peak period, and developing effective building codes (Li et al. 2013; Edwards, New, and Parker 2012). Researchers have found that building energy systems with accurate energy prediction are expected to save 10% to 30% of total building energy consumption (Colmenar-Santoset al. 2013). Accordingly, efforts spent toward building energy prediction are meaningful and crucial for building energy efficiency.

In the past three decades, many researches and studies had been conducted to develop effective building energy prediction tools. Based on applied algorithms and required data, these tools can be further classified into three main categories: engineering methods based, Artificial Intelligence (AI) based, and hybrid methods based (Foucquieret al. 2013). The engineering method uses physical principles to calculate thermal dynamics and energy behaviors for each building component or on the whole building level. This method is also named as "white-box" method because the inner logic is known. Different from engineering method, the AI-based method is considered a "black-box" method as it investigates building energy usage without investigating each building component's internal contribution to overall building energy usage. Hybrid method, also known as "grey method," integrates both white-box and black-box methods for the purpose of eliminating the limitations, as well as utilizing the advantages, of black and white method. Both the white-box and grey-box methods require detailed building information as their inputs in order to simulate the inner relation between each building component. However, it is difficult and sometimes impossible to acquire this information for existing buildings. What's more, to construct building energy models is time consuming and requires tedious expert work, making it hard to be widely applied. Last but not least, engineering based models require comprehensive and deep information of the building it that is being targeted. This information is extremely difficult to obtain accurately. On the contrary, AI-based building energy prediction methods predict building energy usage according to its correlated variables such as environmental conditions and occupancy status. Since the input information is easy to acquire and the calculations are fast and efficient, AI-based methods have been widely applied by many researchers in the domain of predicting building energy usage.

This paper introduces an advanced machine learning technique, namely homogeneous ensemble model, to the field of building energy prediction. Homogeneous ensemble model improves the predictive ability of traditional single model by generating multiple base models and combining their prediction results. During the ensemble process, the model eliminates many of the inherent limitation within each base model. This method has been successfully applied in many fields, i.e., cancer classification (Tan and Gilbert 2003), highway crash prediction (Saha, Alluri, and Gan 2015), and drug failure prediction (Kang, et al. 2015). Yet the homogeneous ensemble techniques have not been used as a prediction tool in the field of building energy prediction. The aim of this study is to offer the theory, development and application of homogeneous ensemble model to researchers in the field of building science. A typical homogeneous ensemble model called ensemble bagging trees (EBT) is used to validate the feasibility and quality of homogeneous ensemble model in building energy prediction. A case study using the electricity usage data of an institutional building is conducted to test the proposed method. To demonstrate the performance difference between a homogeneous ensemble model and a single model in building energy prediction, a comparison in predictive performance is provided.

The remainder of this paper is organized as follows: section 2 reviews previous work; section 3 discusses the research methodology, followed by an introduction of data collection in section 4; in section 5, the model development as well as the prediction results are presented; the conclusion is drawn in section 6.

Literature Review

Based on the prediction scheme, AI-based prediction methods can be classified into two categories: single and ensemble prediction methods. The single prediction method only has one learning algorithm and one prediction model throughout the prediction process, while the ensemble prediction method contains multiple prediction models and a model structuring and organizing process when outputting the data.

Traditionally, AI-based prediction is conducted based on the single prediction methods and develop one prediction model throughout the prediction process. Various learning algorithms including multiple linear regression (Catalina, Virgone, and Blanco 2008), artificial neural network (ANN) (Platon, Dehkordi, and Martel 2015), and support vector machine (SVM) (Li et al. 2009) have been introduced by researchers in the past two decades and received promising prediction results. Notably, most previous researches were conducted on a "single prediction" model basis that uses one learning model in the model training and prediction process. The single prediction model is easy and efficient to be implemented, however, it is worth mentioning that it has several limitations. Since each of the learning algorithms has its own limitations and, depending on the application and selection of initial parameters and their values, it can be claimed that none of them is superior to all others. Each algorithm has its own problem preferences where it can outperform all others. But none of them consistently outperform all of the others in all the problems in the building energy prediction field. Hence, researchers need to select an appropriate learning algorithm each time for the specific problem. Currently the algorithm selection process is carried out by a heuristic method that requires researchers to have adequate experience. Moreover, the training data may not be sufficient for researchers to confidently find the best learning algorithm. Another major issue is the instability of learning algorithms where its performance is highly dependent on the starting value of tuning parameters. Inappropriate selection of the initial parameter values could result in the algorithm getting stuck in local optimal solution instead of overall optimal. Previous researches indicated that some learning algorithms such as ANN and RT are unstable (Breiman 1994). The prediction model may result in major errors in the output because of its sensitivity to the small changes in the training data.

To overcome the limitation as well as to improve the prediction performance, one approach is to combine these single models in either a sequential or parallel manner to cancel out the inherent errors found in each of these models. This approach is generally called "ensemble learning." Through the seamless linkages to multiple AI models, ensemble learning models have exhibited higher accuracy in several research domains, particularly in data classification (Giacinto and Roli 2001), disease diagnosis (Tan and Gilbert 2003), and power system load predictions (Siwek, Osowski , and Szupiluk 2009). Yet, this approach is not prevalent in the building- related field, and the related research was not started until 2014.

Fan et al. (Fan, Xiao, and Wang 2014) initially developed a data mining based ensemble model to predict next-day energy consumption and peak power demand of a commercial building. This ensemble model contains eight base models which were trained independently by different prediction algorithms. The generic algorithm (GA) was used to combine the base models and output the final results. The research indicates that the prediction accuracy of the proposed ensemble model is higher than those of the individual models.

Chou and Bui (Chou and Bui 2014) used AI models individually and in combination (ensemble models) to predict residential building cooling load (CL) and heating load (HL) in the building design stage. Twelve building types simulated in Ecotect were investigated in their research. All buildings had the same volume and the same materials but had different surface areas and dimensions. Activities in the buildings were assumed to be sedentary. Various data mining techniques, including SVR, ANN, classification and regression tree (CART), chi-squared automatic interaction detector, general linear regression, and ensemble inference model were used as base model for prediction. Eight building characteristics were used as the input to predict CL and HL. Comparison results indicated that the ensemble model (SVR + ANN) and SVR were found to be the best models for predict CL and HL, respectively. Their research supports the feasibility of using ensemble model to facilitate early designs of energy conservative buildings.

More recently, Jovanović et al. (Jovanović, Sretenović, and Živković 2015) used a neural network- based ensemble model to predict daily heating energy consumption. Three artificial neural networks, i.e., feed forward neural network (FFNN), radial basis function neural network (RBFN), and adaptive network-based fuzzy inference system (ANFIS) were used to build the ensemble model. Three different methods namely the simple average, weighted average, and median based averaging for combining base models were used. The results showed that all proposed neural networks were able to predict heating consumption with great accuracy, and that using ensemble achieves even better results.

Notably, ensemble prediction methods can be classified as heterogeneous or homogeneous based on the base model generation method. Heterogeneous ensemble models generate base models by training different models using different learning algorithms or parameter settings, while homogeneous ensemble models generate base models by training the model using the same learning algorithm with different subsets of the training data. Based on the reviews, the heterogeneous ensemble model has been limitedly introduced in building energy prediction, while the application of homogeneous ensemble model in building science, particularly in building energy prediction, is underexplored.

Methodology

This section introduces the general framework of homogeneous ensemble model, the theory of EBT, and the performance evaluation indices.

Homogeneous Ensemble model

The data collection and data preprocessing for ensemble models are similar to single models. However, the major difference between these two types of methods lies in the process of selecting and training the learning algorithms. As opposed to single prediction models that generally contains only one single learning model, an ensemble model consists of multiple learning models, known as base models. The base models are developed by resampling, manipulation or randomization of the training data, learning algorithm, and learning parameters. Based on the generation of the base models, ensemble prediction models can be further classified into two types: the homogeneous and the heterogeneous ensemble models. The homogeneous ensemble model uses the same learning algorithm on different distributions of the training set, i.e., bagging and boosting (Reid 2007). This method works especially well for unstable learning algorithms – algorithms whose output undergoes major changes in response to small changes in the training data (Dietterich 1997).

Figure 1 depicts the general framework of homogeneous ensemble model. After the training data is obtained, a training data subsampling process is introduced to generate multiple subsets of the training data. The training data subsets are then used to train the same learning algorithm to generate multiple base models. Notably, the base models can be generated either simultaneously or iteratively, depending on the applied base model generation method. Finally, a base model combination process is conducted to integrate all base models to output the final prediction results.

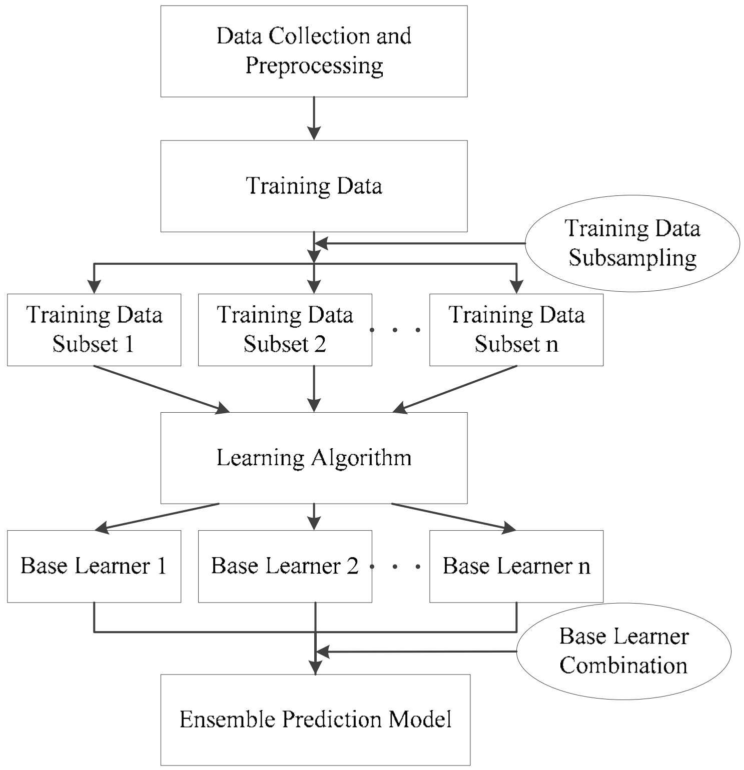

Figure 1. General framework for homogeneous ensemble model.

Ensemble Bagging Trees

Ensemble bagging trees (EBT) generates multiple RT based base models by repeatedly resampling training data with replacement, and combining the result of each RT trees to generate the overall result (Breiman 1996). The model integrates bagging and RT techniques to provide improved prediction results.

Bagging, also named as bootstrap aggregation, was the most straightforward way of manipulating the train data set. Given an original training set D of size n, bagging generates m new training sets D_i, each of size n, by sampling from D uniformly and with replacement. Each new training set is called a bootstrap replicate of the original training set. By sampling with replacement, some observations may be repeated in each D_i. The new training sets contain, on average, 63.2% of the original training set, with the rest being duplicates (Breiman 1996). Bagging is normally used to improve the performance of unstable learning algorithms. An instability study of learning algorithms was conducted by Breiman where it was pointed out that neural nets, CART, and subset selection in linear regression were unstable (Breiman 1994).

Classification and regression trees (CART) develop prediction model by recursively partitioning the data space and fitting a simple prediction model within each data space. While the major differences between classification tree (CT) and RT are the type of target variables and the way of measure prediction error. CT is used to predict target variables that have a finite set of values, and the prediction error is measured in term of misclassification rate. RT is used to predict target variables that have continuous or ordered discrete values, and the prediction error is measured by the squared error between the predicted and actual values.

Performance Evaluation Indices

To evaluate the prediction performance of the models, the following evaluating indicies are computed:

Coefficient of determination (R^2) measures the goodness of fit of a model. A high R^2 value indicates the predicted values perfectly fit the observed values. The R^2 is defined as follows:

$$R^2 = \frac{\sum_{t=1}^{n}(x_t - \bar{y})^2}{\sum_{t=1}^{n}(y_t - \bar{y})^2}$$

where n is the sample size; x_t is the predicted value; y_t is the observed value; and $\bar{y}$ is the mean of y_t.

Root Mean Square Error (RMSE) stands for the sample standard deviation of the differences between predicted and observed values. This measures is used to identify large errors. RMSE amplifies and severely punishes large error. The mathematical formula for RMSE is:

$$RMSE = \sqrt{\frac{1}{n} \times \sum_{t=1}^{n}(x_t - y_t)^2}$$

Mean Absolute Percentage Error (MAPE). The MAPE is a statistical measure to describe the accuracy of the prediction by comparing the residual with the observed values. It usually expresses accuracy in percentage, and is effective for evaluating the performance of the prediction model by introducing the concept of relative values. The MAPE is defined by formula:

$$MAPE = \frac{1}{n} \times \sum_{t=1}^{n}\left|\frac{x_t - y_t}{y_t}\right| \times 100\%$$

Data Collection

The data set adopted in this research contains two major parts: the input data and the output data. The input data include variables which impact building energy performance such as metrological variables and number of occupants. While the output data contains building level electricity usage data, collected at 15-minute intervals from Rinker Hall. Rinker Hall is a three-level building within the University of Florida's College of Design, Construction, and Planning. The building has a floor area of 47,270 ft^2 and includes a mix of classrooms, offices, laboratories, and student facilities. It is designed to accommodate 38 permanent occupants and up to 450 students. Completed in 2003, Rinker Hall was the first building in Florida to be designed under the LEED program and finally received LEED Gold certification in 2004.

Table 1 lists independent variables used in this research. The weather station operated by Department of Physics at University of Florida's main campus is the primary source of information for metrological data. Metrological data including outdoor temperature, dew point, relative humidity, barometric pressure, precipitation, wind speed, and solar radiation are extracted in hourly average basis from the weather station's historical database. Three variables that describe the time of observations, i.e. workday type, day type, and time of day, are considered in this research as they reveal weekly, daily, and hourly occupants' working patterns. In addition, the number of occupants for each hour is estimated based on the building's daily working and class schedule which is obtained from the university's Space Inventory & Allocation System. Figure 1 shows Rinker Hall hourly based number of occupants which is calculated based on the building's class schedule for 2015 spring semester.

The data collection period lasts for 4 weeks (2015.02.01-2015.02.28), which covers a total number of 28 calendar days. Consequently, a total number of 672 (28*24) data points were collected.

Table 1. List of independent variables used for prediction

Variable	Type	Unit
Outdoor Temperature	Continuous	F
Dew Point	Continuous	F
Relative Humidity	Continuous	%
Barometric Pressure	Continuous	Inch
Precipitation	Continuous	inHg
Wind Speed	Continuous	mph
Solar Radiation	Continuous	W/m2
Workday Type	Categorical	weekday and weekend
Day Type	Categorical	Sunday, Monday, …, Saturday
Time of Day	Categorical	1, 2, 3,…, 23, 24
Number of Occupants	Continuous	person

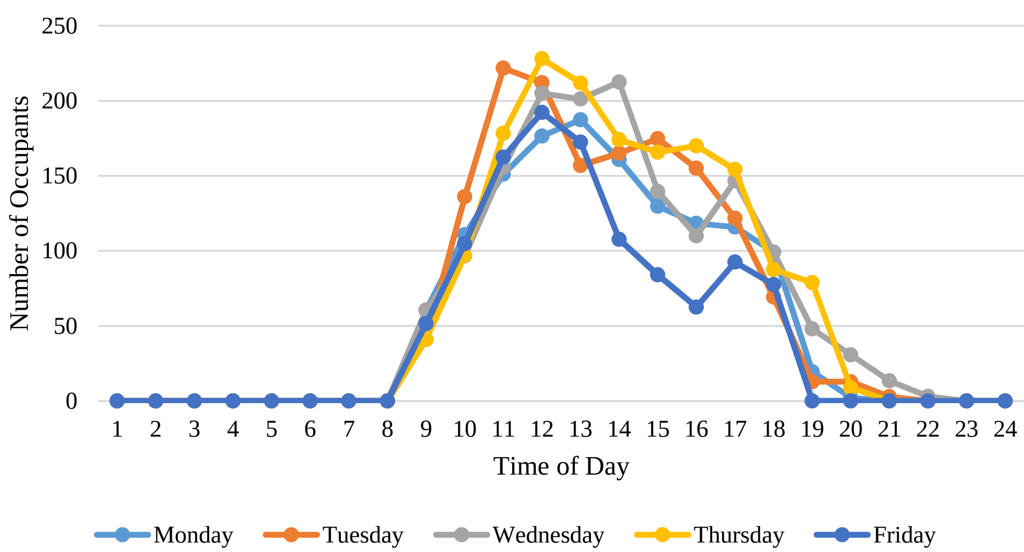

Figure 2. Hourly based number of occupants for Rinker Hall (2015 spring semester).

Model Development and Results

The obtained data are randomly divided into two parts: the training data and the testing data, for the purpose of training and testing the proposed prediction models. As a result, 600 data points (89%) are used as training data and the rest 72 data points (11%) are used as testing data. MATLAB R2015b is used as the programming language to develop all prediction models.

To develop the EBT model, two important parameters, namely the leaf size and the number of trees, are used to tune the model. The leaf size determines the predictive ability of each RT based base model within the ensemble model. The number of trees determines the size and the overall performance of the ensemble model.

Optimal Leaf Selection

The leaf size of the RT is critical to the speed and quality of the learning process. The leaf size determines how deep the tree will grow. A small leaf size results in a deep tree that has good generalization ability although it requires more training and prediction time as well as memory consumption. On the contrary, a large leaf size can bring a shallow tree that requires less calculation time and memory, however, its accuracy is not guaranteed and is dependent on the complexity and availability of the dataset. In this study, the optimal leaf size is selected based on average out-of-bag error of EBT. As we mentioned in section 2, bagging technique generates new training data subsets by randomly kicking out on average 37% of the initial training data and replace the data with a duplication from the rest of the 63% of the data. The omitted training data for each new training data subset, which are known as out-of-bag data, can be used to measure the predictive performance of the generated RT based base models. The average out-of-bag error is calculated by averaging the out-of-bag data's mean square prediction error of each RT based base model within the EBT. In this study, five typical leaf sizes, i.e., 5, 10, 20, 50, and 100 are used to select the optimal leaf size. Figure 3 shows the average out-of-bag errors of the ensemble tree for different leaf sizes. It can be seen from figure 3 that the red curve (leaf size 5) has the lowest average out-of-bag error. Hence, the optimal leaf size is selected as 5. This result matches our expectation as 5 will offer the deepest tree growth. Since our dataset is not excessively huge, the computational cost in terms of time and complexity is controlled within acceptable range.

Selection of Number of Trees

After the optimal leaf size is selected, the next step is to select an appropriate number of trees for the EBT. In general, the EBT should have sufficient base models in order to eliminate the inherent limitation of these models. However, a large ensemble would cost too much time to develop the model and make prediction. In this study, the number of trees are determined by trial and error. A stepwise validation ranges from 1 to 300 with a step width of 1 is applied to select the optimal number of trees. Figure 4 shows the average out-of-bag error of ensembles with different number of trees. It is easy to observe that the average out-of-bag error curve converges after the ensemble model has more than 100 trees. To reduce unnecessary computational cost, the number of trees is set to 100.

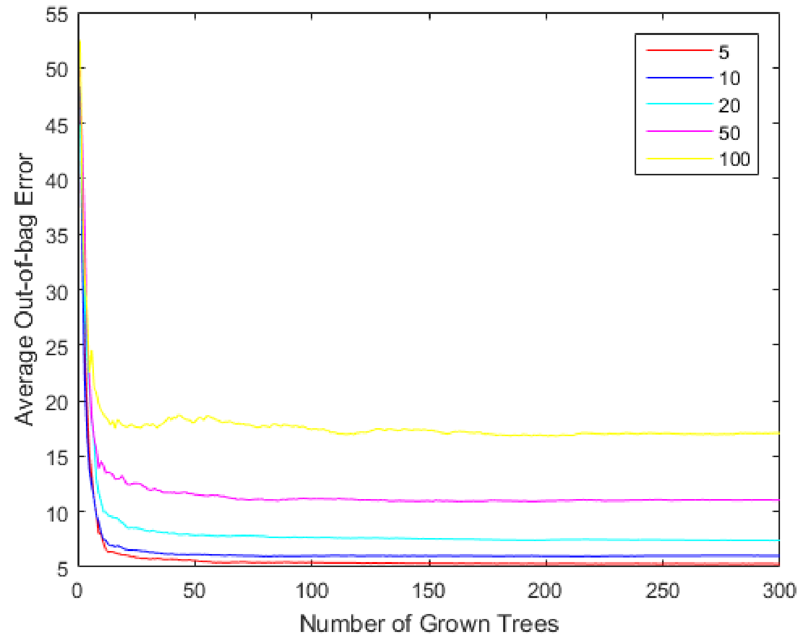

Figure 3. Average out-of-bag error for different leaf sizes.

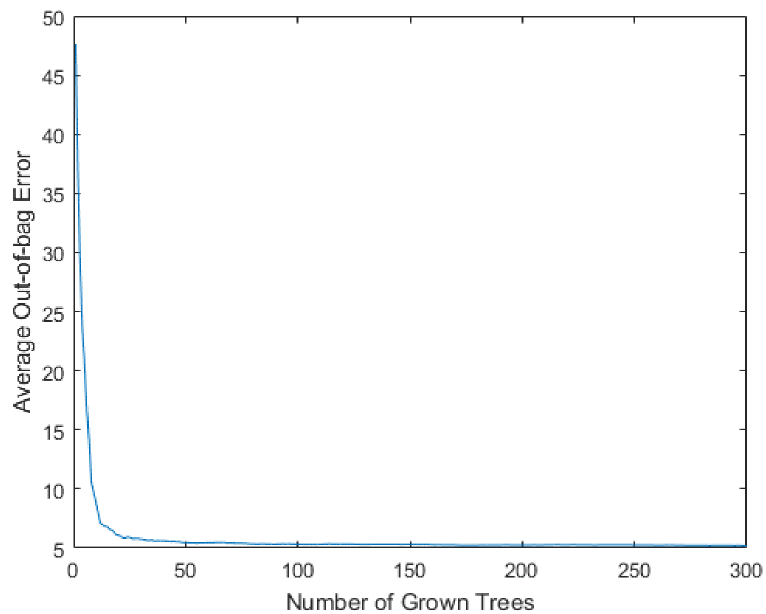

Figure 4. Average out-of-error for different number of trees.

Prediction Results

In this study, a conventional RT is set as the baseline model in order to compare the predictive performance between the single and ensemble model. To receive reliable results, the comparison was repeated 10 times. Figure 5 shows an example of testing results. Table 2 summarize some statistic properties including the CV, RMSE, and MAPE values of these 10 comparisons between RT and EBT. The results suggest that both models were highly predictive based on the range of R^2 values from 0.82 to 0.95. The promising predictions from both models are likely due to the strong dependence of building electricity usage on the selected environmental and occupancy variables. However, the descriptive statistics indicate EBT performs better than RT in predictive accuracy, based on the average values of R^2 (0.93 vs. 0.88, respectively), RMSE (2.13 vs. 2.83, respectively), and MAPE (3.17% vs. 3.89%, respectively). Meanwhile, to evaluate the instability of the prediction models, the standard deviations of the performance evaluation indices are calculated. It can be found in table 2 that EBT has lower standard deviations of CV, RMSE than RT while the standard deviation of MAPE for these two models are similar. This also matches the expectation because the ensemble model eliminates the inherent limitation of each base model and utilizes the advantages of each base model. It makes sense that the ensemble model could outperform any single base model. The results suggest that EBT are more stable than RT in predictive ability.

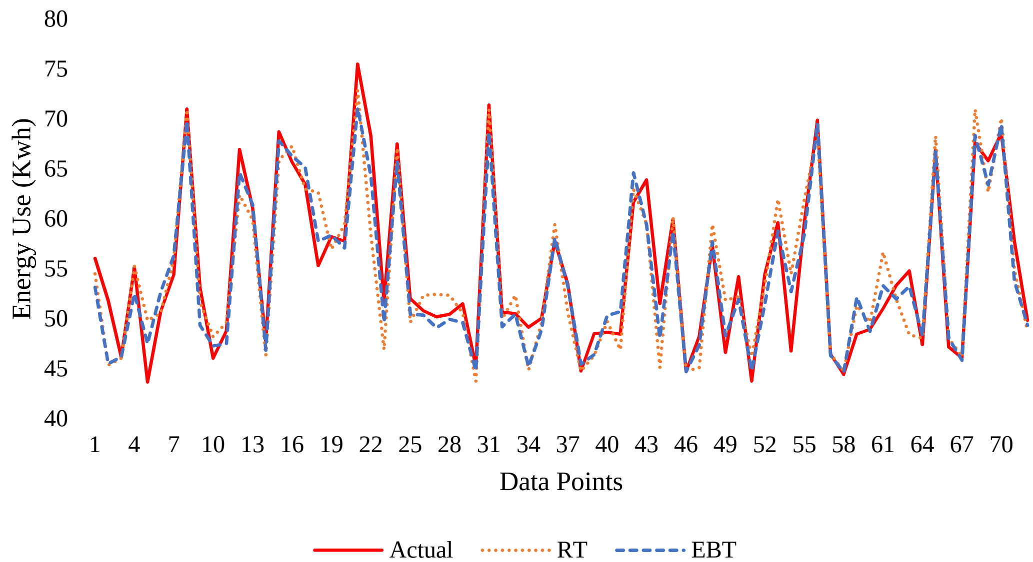

Figure 5. Testing results.

Table 2. Summary statistics of the predictive quality of regression tree (RT) and ensemble bagging tree (EBT).

Model	Index	Min	Mean	Max	Standard Deviation
RT	R^2	0.82	0.88	0.93	0.029
	RMSE	2.10	2.83	3.39	0.335
	MAPE	3.15%	3.89%	4.42%	0.003
EBT	R^2	0.90	0.93	0.95	0.016
	RMSE	1.86	2.13	2.47	0.189
	MAPE	3.15%	3.17%	3.61%	0.003

Conclusion

This study proposed a homogeneous ensemble model namely, ensemble bagging tree (EBT) for building energy prediction. An institutional building located in the University of Florida was selected as the target building for energy prediction. Training and testing data were collected through various University of Florida facilities. The general framework of homogeneous ensemble model as well as the theory of EBT were presented. To demonstrate the feasibility and quality of EBT in building energy prediction, a case study was conducted.

The experiment was held in Rinker Hall and eleven variables including both environmental and occupancy information were used as the input data of the prediction model. The output data are hourly building level electricity usage data. A comparison between EBT and RT was conducted to investigate the difference in prediction accuracy and stability between these two methods. The results indicated that EBT provides better predictive performance than RT, with an average of 18.5% improvement in MAPE. In addition, the instability analysis showed EBT performs more stable than RT.

Through the use of data collected from the University of Florida facilities, further investigation will focus on selecting the best base model training algorithm for homogeneous ensemble models as well as comparing heterogeneous with homogeneous ensemble models in solving the same problem.

References

Breiman, L. 1994. *Heuristics of instability in model selection*. University of California at Berkeley.

Breiman, L. 1996. "Bagging predictors." *Machine Learning* 24(2): 123-140.

Catalina, T., Virgone, J., & Blanco, E. 2008. "Development and validation of regression models to predict monthly heating demand for residential buildings." *Energy and Buildings* 40(10): 1825–1832.

Chou, J.-S., & Bui, D.-K. 2014. "Modeling heating and cooling loads by artificial intelligence for energy-efficient building design." *Energy and Buildings* 82: 437-447.

Colmenar-Santos, A., Terán de Lober, L., Borge-Diez, D., & Castro-Gil, M. 2013. "Solutions to reduce energy consumption in the management of large buildings." *Energy and Buildings* 56: 66-77.

Dietterich, T. 1997. "Machine learning research: Four current direction." *AI Magazine* 18(4): 97-136.

Edenhofer, O. (2015). King Coal and the queen of subsidies. Science, 349(6254), 1286-1287.

Edwards, R. E., New, J., & Parker, L. E. 2012. "Predicting future hourly residential electrical consumption: A machine learning case study." *Energy and Buildings* 49: 591-603.

Energy Information Administration. 2015. *The annual Energy Outlook 2015*. Energy Information Administration.

Fan, C., Xiao, F., & Wang, S. 2014. "Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques." *Applied Energy* 127: 1–10.

Foucquier, A., Robert, S., Suard, F., Stéphan, L., & Jay, A. 2013. "State of the art in building modelling and energy performances prediction: A review." *Renewable and Sustainable Energy Reviews* 23: 272-288.

Giacinto, G., & Roli, F. 2001. "Design of effective neural network ensembles for image classification purposes." *Image and Vision Computing* 19: 699-707.

Jovanović, R., Sretenović, A., & Živković, B. 2015. "Ensemble of various neural networks for prediction of heating energy consumption." *Energy and Buildings* 94: 189-199.

Kang, S., Kang, P., Ko, T., Cho, S., Rhee, S.-j., & Yu, K.-S. 2015. "An efficient and effective ensemble of support vector machines for anti-diabetic drug failure prediction." *Expert Systems with Applications* 42(9): 4265–4273.

Li, N., Kwak, J.-y., Becerik-Gerber, B., & Tambe, M. 2013. "Predicting HVAC Energy Consumption in Commercial Buildings Using Multiagent Systems." *International Symposium on Automation and Robotics in Construction*. Montreal.

Li, Q., Meng, Q., Cai, J., Yoshino, H., & Mochida, A. 2009. "Applying support vector machine to predict hourly cooling load in the building." *Applied Energy* 86(10): 2249-2256.

Platon, R., Dehkordi, V. R., & Martel, J. 2015. "Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis." *Energy and Buildings* 92: 10–18.

Reid, S. 2007. *A Review of Heterogeneous Ensemble Methods.* University of Colorado at Boulder, Department of Computer Science. Retrieved from http://buildingsdatabook.eren.doe.gov/

Saha, D., Alluri, P., & Gan, A. 2015. "Prioritizing Highway Safety Manual’s crash prediction variables using boosted regression trees." *Accident Analysis & Prevention* 79: 133-144.

Siwek, K., Osowski , S., & Szupiluk, R. 2009. "Ensemble Neural Network Approach for Accurate Load Forecasting in a Power System." *International Journal of Applied Mathematics and Computer Science* 19(2): 303-315 .

Tan, A., & Gilbert, D. 2003. "Ensemble machine learning on gene expression data for cancer classification." *Proceedings of New Zealand Bioinformatics Conference*, (pp. 75-83). Te Papa.

U.S. Energy Information Adminstration. (2016, May 09). *International Energy Statistics*. Retrieved from Independent Statistics & Analysis: http://www.eia.gov/cfapps/ipdbproject/IEDIndex3.cfm?tid=44&pid=44&aid=2