

Prediction of Hourly Power Consumption for a Central Air-Conditioning System Based on Different Machine Learning Methods

Si-qi Gao^{1,2,3}, Fu-min Zou^{2(✉)}, Xin-hua Jiang³, Lyuchao Liao³, and Yun Chen³

¹ Beidou Navigation and Smart Traffic Innovation Center of Fujian Province Fuzhou, Fuzhou 350118, Fujian, China

784912077@qq.com

² Fujian University of Technology, Fuzhou 350118, Fujian, China

fmzou@fjut.edu.cn

³ Fujian Key Laboratory for Automotive Electronics and Electric Drive, Fujian University of Technology, Fuzhou 350118, Fujian, China

achao@fjut.edu.cn

Abstract. This paper uses a variety of machine learning methods to predict the hourly power consumption of a central air-conditioning system in a public building. It is found that the parameters of the central air-conditioning system are different at different times, so is the corresponding power consumption. The paper applies the time series to predict the power consumption on account of the time, which predicts the hourly power consumption based on historical time series data. Comparing the prediction accuracy of multiple machine learning methods, we find that the Gradient Boosting Regression Tree (GBRT), one of the ensemble learning methods, has the highest prediction accuracy.

Keywords: Central air-conditioning system · Machine learning methods · Ensemble learning · Time series prediction

1 Introduction

With the information technology shooting up, intelligent buildings and energy issues have become the focus of increasing attention. Central air conditioning, as an indispensable part of intelligent building, often accounts for a large proportion of the energy consumption in large public buildings. Energy is largely lost in that the system works at low temperature and large flow rate for a long time in the case of low load and low temperature. Therefore, if the system power consumption of each moment can be accurately predicted, the optimization plan could be designed in advance to reduce the energy consumption.

Conventional prediction models are based on physical principles to calculate thermal dynamics and energy behavior at the building level. Some of them include models of space systems, natural ventilation, air conditioning system,

passive solar, photovoltaic systems, financial issue, occupants behavior, climate environment, and so on [1]. There are also statistical methods [2–4]. These methods are used to predict energy consumption and that affecting variables such as weather and energy cost. In addition, in order to influence the development of future building systems, there are also some hybrid methods [5,6], which combine the models above to optimize performance. Conventional methods are advantages in that they are easily implemented and interpreted, however, their disability to handle non-linearity in short-term load series instead encourages the use of machine learning methods [7]. In recent years, a large number of machine learning algorithms have been used to predict energy consumption [8–10].

The paper aims to predict the system hourly power consumption based on historical data of the running parameters of the various system components, which utilizes the four machine learning algorithms: Random Forests, K-Nearest Neighbour, Gradient Boosting Tree Regression and Support Vector Regression. Moreover, seeking a model that is easy to be implemented with minimal input requirements and maximum accuracy.

2 Data Exploratory Analysis

The data source used in the paper has 51 variables (including timestamp). There are totally 1750 instances (from October 4, 2016 at 10:00 to December 29, at 13:00 once every hour to collect) that is an actual data set collected from sensors of central air conditioning system parameters. The following Table 1 shows the partial variables description and the remaining variables are specified in the paper.

Table 1. Partial variables description

Variables	Description	Unit
chwsfhdr	The velocity difference of the cooler	gal/min
$ch_i kw (i = 1, 2, 3)$	The power consumption of the cooling device	kw
chwrhdr	The water temperature that flows into the cooler	°C
rh	Relative humidity	%
drybulb	Drybulb temperature	°C
effsys	Effect of the system	kw/RT
systotpower	Total power consumption of the system	kw
loadsys	Load of the system	RT
cwgpmt	Speed of load flow	gal/min·RT
ct_pc	Speed of cooling tower fan (25 Hz–50 Hz)	Hz

2.1 Data Preprocessing

Before digging into the characteristics and regularities of the data changes, we first utilize the data preprocessing, which consists of data cleaning, data integration and data reduction.

Data Cleaning. According to the principle of heat balance of the system, we filter out the data points when the system is stable (when the heat balance $> 5\%$ means the system is unstable). Then the system power is drawn by time series and it is found that there are several missing values. Here we directly eliminate in that the small amount and the final effect is shown in Fig. 1.

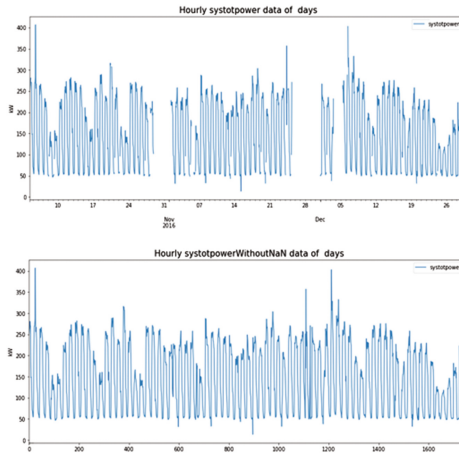


Fig. 1. Time series diagram of hourly system power consumption

Data Integration and Data Reduction. Since the data set of the central air conditioning system has many attributes, in view of the integration of each other, the correlation coefficients between attributes are obtained, and the irrelevant attributes are removed by dimension reduction, thus reducing the amount of data mining.

The Pearson correlation coefficient reflects the correlation between two variables. The closer the absolute value is closer to 1, the closer the correlation between the operating parameters. Similarly, the closer the absolute value is closer to 0, the lower the correlation is. After analysis of the data integration and the Pearson correlation coefficient between the attributes and systotpower, operating parameters that are significantly related (the absolute value of Pearson correlation coefficient greater than 0.5) to systotpower are selected. A total of 24 variables were eliminated and 26 variables remained.

3 Mathematical Background

3.1 Ensemble Learning Methods (Random Forests and GBRT)

The core of ensemble learning is to build multiple different models and aggregate these models to improve the performance of the final model. According to the generation strategy of base learner, ensemble learning can be divided into two categories: (1) parallel methods, take the bagging as the main representative; (2) sequential methods, take the boosting as the main representative.

Random Forests. Random Forests is a typical application algorithm of bagging. Bagging is abbreviated as Bootstrap AGGREGating, that is, a number of different training sets are constructed by means of bootstrap sampling (repeatable sampling). Then the corresponding base learner is trained on each training set. Finally, these base learners are aggregated to obtain the final model. Compared with a single model, the disadvantages of ensemble learning are: (1) computational complexity, in that multiple models need to be trained. (2) the resulting model is difficult to interpret.

However, the random forest can estimate importance of the variable. Therefore, a random forest model directly tells us the importance of each variable, which is beneficial for eliminating irrelevant variables and redundant data. Not only could it improve the performance of the model but also reduce the computational complexity.

Gradient Boosting Regression Tree. Gradient boosting algorithm was fathered by FreidMan in 2000, which is another method of ensemble learning methods. The core is that each tree learns from the residuals of all previous trees. The negative gradient of the loss function $L(y_i, f(x_i))$ is used in the current model as the approximate value of the residuals in the boosting algorithm to fit a regression (classification) tree. The negative gradient of the loss function has the form:

$$r_{mi} = -\left[\frac{\partial L(y_i, f(x_i))}{\partial f(x_i)}\right]_{f(x)=f_{m-1}(x)}.$$

For the Gradient Boosting algorithm, multiple learners are sequentially build. Using the correlation of the base learner, the existing shortcomings of weak learners is overcome, which the characterization is the gradient. Gradient Boosting selects the descent direction of gradient during iteration to improve the performance after polymerization. The loss function is used to describe the reliable degree of the model. If the model is not over fitting, the greater the loss function, the higher the error rate of the model. If our model allows the loss function to continue to decline, it indicates that our model is constantly improving, and the best way is to let the loss function fall in its gradient direction.

3.2 K-Nearest Neighbor Regression

K-Nearest Neighbor is a simple algorithm that stores all available cases and predicts digital targets based on similarity (e.g. distance function). KNN, as a non-parametric technique, has been used in statistical estimation and pattern recognition in early 1970. KNN regression is a non-parametric learner, namely, you do not need to give a function such as linear regression and logistic regression, and then to fit parameters. Non-parametric learner is a function without guesswork, and it eventually fail to get an equation, but it could acquire fine prediction. The advantage of non-parametric learner is that it's easy to change the model and the training speed is fast. The disadvantage is that all the points need to be stored so that the space is consumed and the query is slow.

3.3 Support Vector Regression

Support vector regression (SVR) is one of the most popular algorithms for machine learning and data mining. SVR is derived from support vector machines (SVM) and mainly used to fit the sample data and predict the unknown data.

The input data is first mapped to a high-dimensional feature space by kernel functions, and then the linear regression functions in high-dimensional feature space are computed. In contrast to the expression of both Linear Regression and SVR, the latter is a constrained optimization problem that allows the model to find a tube instead of a line. For the ε -SVR (Epsilon insensitive support vector regression), our goal is to find a function, such as linear function $f(x) = W^T x + b$, to minimize generalization errors:

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + c \sum_{i=1}^l (\xi_i + \xi_i^*) \\ \text{s.t.} \quad & |y_i - W^T x_i - b| < \varepsilon + \xi_i + \xi_i^* \\ & \xi_i, \xi_i^* \geq 0 \end{aligned}$$

A loss function is introduced here, which is used to ignore the fluctuation error in the range of the predicted value and the true value, namely, if the deviation of the data from the regression function subjects to the constraint: $|y_i - W^T x_i - b| < \varepsilon$, the data can be ignored. This constraint guarantees the fitting of the best linear regression function so that more points fall within the range of accuracy accepted. But it can be found that there are still some points whose deviations are quite large, so relaxation factors ξ, ξ_i^* are introduced. SVR is suitable for solving small sample, nonlinear and high-dimensional problems, and also has great potential to overcome overfitting and curse of dimensionality.

4 Experiments and Results

In order to achieve the goal of hourly power consumption prediction, four machine learning models were evaluated and compared with a set of measured

Table 2. Feature importance

Variables	Importance	Variables	Importance	Variables	Importance	Variables	Importance	Variables	Importance
chwsfhdr	0.980011	ch1kw	0.000740	ct1kw	0.000264	loadsys	0.000187	ch1stat	0.000097
chwrhdr	0.006654	ch2kw	0.000528	ct2kw	0.000262	cwp_pc	0.000163	chwp1stat	0.000094
cwshdr	0.004497	cwp1kw	0.000372	chwp4kw	0.000257	chwp_pc	0.000126	cwp1stat	0.000085
cwrhdr	0.003368	cwp2kw	0.000324	drybulb	0.000199	ct_pc	0.000122	cwp3stat	0.000025
cwsfhdr	0.001028	cwp3kw	0.000288	chwp1kw	0.000190	dch	0.000104	ct2stat	0.000017

data. In the data set, operating parameters collected show the evolution of the overall system power over time in a central air-conditioning system. The data set contains 1750 instances, nearly 3 months’ data, which is collected hourly. Instances from October 4, 2016 at 10:00 to December 1, 2016 at 13:00, 1127 time series data in total, are as the train set, 623 instances remained as the test set (Table 2).

4.1 Experiments Details

We have implemented data fitting and prediction models in Python2.7 using four machine learning algorithms described in Sect. 3. We first trained a random forest containing 1000 decision trees to assess the importance of 25 dimensional features preprocessed in Sect. 2, results are shown in Table 3.

From the result analysis, we can see that the state of air conditioning water system is the independent variable, so we can eliminate these 5 variables: ch1stat, chwp1stat, cwp1stat, cwp3stat, ct2stat. The remaining 20 variables in Table 4 are used for learning of the latter three models.

Table 3. Variables used in the training model

chwrhdr	cwshdr	loadsys	chwp_pc	ct1kw	ch1stat	chwsfhdr
ch1kw	cwsfhdr	chwp4kw	chwp1stat	ch2kw	cwp3kw	cwp_pc
cwp1stat	cwp2kw	cwp3stat	cwrhdr	cwp1kw	drybulb	dch
ct2kw	ct_pc	chwp1kw	systotpower	ct2stat		

Table 4. MSE and R^2 score of the four models

Model	MSE	R^2	
		Train set	Test set
RF	32.33	0.99934	0.99452
GBRT	4.49	0.99999	0.99925
KNN	253.70	0.98265	0.95697
SVR	7.14	0.99513	0.99879

4.2 Prediction of Hourly Power Consumption

Predicting the power consumption of office buildings will be directly affected by human behavior and other factors. All these factors lead to a nonlinear time series [1]. The estimation and prediction accuracy of the four models used in this paper show good results, Table 4 and Figs. 2, 3, 4, 5, 6, and 7 give detailed results for all models.

In this paper, we use combination of grid search and cross validation to seek the k with the highest R^2 score from $\{1, 2, 3, 4, 5, 6, 7, 8, 9, 10\}$ as the optimal parameter in KNN regression model, and then the k is used to prediction. Experiment result shows that cross validation with K -fold ($K = 10$) returns the best estimator and result is shown in Fig. 4. The best $k = 3$ found by 10-fold cv, and the predictive values are plotted with a scatter chart for the next Fig. 5.

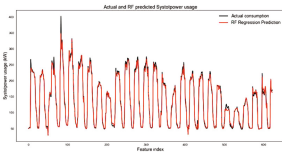


Fig. 2. RF

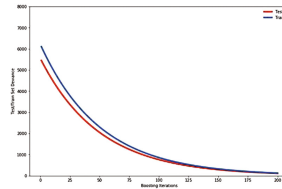


Fig. 3. GBRT

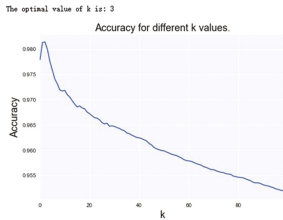


Fig. 4. Seeking the best k

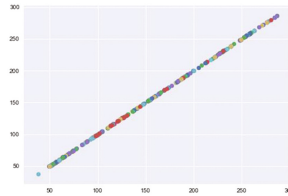


Fig. 5. Scatter plot of predicted systotpower ($k = 3$)

Finally, we will train a SVR model, where the ϵ -SVR is used for regression prediction. During the design, we also used grid search and cross validation to fine tune the SVR parameters. And find the optimal parameters: Optimum parameters $C = 1$ and $\gamma = 0.01$ for SVR. Optimum ϵ and kernel for SVR: $\{\epsilon:0.1, \text{kernel}:'linear'\}$. Figure 7 shows the prediction of hourly systotpower using SVR.

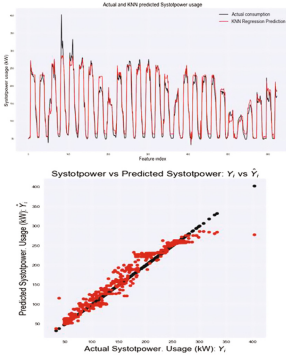


Fig. 6. KNN

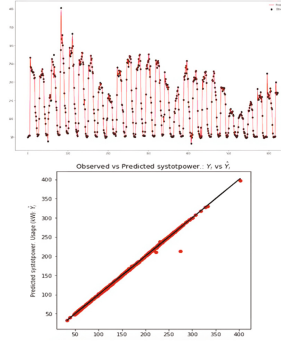


Fig. 7. SVR

5 Conclusion

Our goal is to predict the hourly system power consumption based on historical data and find the optimal model. Experiment result shows that the winner is GBRT. Although the machine learning algorithms are very successful in fitting, when the data is complex and mess, the choose of features and parameters are very important. We have to repeatedly adjust the features and parameters to improve the model.

Acknowledgments. This work is supported by National Natural Science Foundation of China (No. 61304199), Fujian Science and Technology Department (No. 2014H0008), Fujian Transportation Department (No. 2015Y0008), Fujian Education Department (No. JK2014033, JA14209, JA1532), and Fujian University of Technology (No. GYZ13125, 61304199, GY-Z160064). Many thanks to the anonymous reviewers, whose insightful comments made this a better paper.

References

1. Mocanu, E., Nguyen, P.H., Gibescu, M., et al.: Comparison of machine learning methods for estimating energy consumption in buildings. In: International Conference on Probabilistic Methods Applied To Power Systems, pp. 1–6. IEEE (2014)
2. Williams, K.T., Gomez, J.D.: Predicting future monthly residential energy consumption using building characteristics and climate data: a statistical learning approach. *Energy Buildings*. **128**, 1–11 (2016)
3. Braun, M.R., Altan, H., Beck, S.B.M.: Using regression analysis to predict the future energy consumption of a supermarket in the UK. *Appl. Energy* **130**(5), 305–313 (2014)
4. Takeda, H., Tamura, Y., Sato, S.: Using the ensemble Kalman filter for electricity load forecasting and analysis. *Energy* **104**, 184–198 (2016)
5. Li, X., Ding, L., Lv, J., et al.: A novel hybrid approach of KPCA and SVM for building cooling load prediction. In: International Conference on Knowledge Discovery and Data Mining, pp. 522–526. IEEE (2010)

6. Vonk, B.M.J., Nguyen, P.H., Grand, M.O.W., et al.: Improving short-term load forecasting for a local energy storage system. In: Universities Power Engineering Conference, pp. 1–6. IEEE (2012)
7. Hedn, W.: Predicting hourly residential energy consumption using random forest and support vector regression? An analysis of the impact of household clustering on the performance accuracy, Dissertation (2016)
8. Abdelkader, S.S., Grolinger, K., Capretz, M.A.M.: Predicting energy demand peak using M5 model trees. In: IEEE, International Conference on Machine Learning and Applications, pp. 509–514. IEEE (2016)
9. Liu, D., Chen, Q., Mori, K.: Time series forecasting method of building energy consumption using support vector regression. In: IEEE International Conference on Information and Automation, pp. 1628–1632. IEEE (2015)
10. Chen, Y., Tan, H.: Short-term prediction of electric demand in building sector via hybrid support vector regression. *Appl. Energy* (2017)
11. Shuang, H., Linlin, Z., Xiaohua, W.: The study of mechanical state prediction algorithm of high voltage circuit breaker based on support vector machine. *High Voltage Apparatus* **7**, 155–159 (2015)