Sustainable Cities and Society 51 (2019) 101698

Contents lists available at ScienceDirect

Sustainable Cities and Society

journal homepage: www.elsevier.com/locate/scs

A simplified HVAC energy prediction method based on degree-day

Huajing Sha[a], Peng Xu[a,*], Chonghe Hu[b], Zhiling Li[b], Yongbao Chen[a], Zhe Chen[a]

[a] Department of Mechanical and Energy Engineering, Tongji University, Shanghai, 201804, China
[b] Shanghai Yanhua Smartech Group Co., Ltd, Shanghai, 200060, China

ARTICLE INFO

Keywords:
Data-driven model
Energy prediction
Degree-day

ABSTRACT

A building heating, ventilation, and air-conditioning (HVAC) system consumes large amounts of energy. Energy consumption prediction is an effective strategy for operation optimization and energy management in a building. The energy consumption of an HVAC system in a building is influenced by many factors, such as weather conditions, building usage, and thermal performance. However, it is impractical to consider all factors for predicting energy consumption. In this paper, a simplified data-driven model is proposed for predicting the energy consumption of an HVAC system in a building. A novel feature transformation method is introduced to select the most relevant features. Three input features (i.e., degree-day, day type, and month type) are finally adopted in this model. Compared to models developed in previous studies, this simplified model largely reduces the computation time and is easier to operate. The cross-validated root mean square error of this method for cooling energy prediction is less than 20%, indicating its suitability for use in engineering applications.

1. Introduction

The building industry is one of the largest primary energy consumers worldwide, accounting for over 30% of global energy usage (Ürge-Vorsatz, Cabeza, Serrano, Barreneche, & Petrichenko, 2015). In China, heating, ventilation, and air-conditioning (HVAC) systems consume over 40% of the energy used in all building service systems. Related research has stated that most HVAC systems have different levels of energy-saving potential, ranging from 15% to 30% (Pérez-Lombard, Ortiz, & Pout, 2008). Building energy efficiency is of great significance to global sustainability. Thus, techniques such as fault diagnosis (Li, Bowers, & Schnier, 2010) and optimized system control, which help improve system efficiency, are gaining attention. In addition, energy quota management, which refers to setting the amount of energy that can be used over a certain period of time (e.g., day or week), is also an efficient approach to reduce consumption. It has now been adopted by large commercial groups, to improve property management levels. Reliable energy prediction is essential for all of the energy-efficient methods mentioned above. The energy predicted acts as baseline a building would consume under normal condition.

There are two main approaches for energy prediction: physical modeling and data-driven modeling. Physical models rely on explicit thermodynamic rules whose formulas and calculating mechanisms are easy to understand. Thus, physical-based models are also called white-box models. Widely used simulation tools, including EnergyPlus and eQuest, have been developed based on physical models. However, detailed building information is necessary for building physical models. In contrast, data-driven models (also known as black-box models) do not require such detailed data of the building (Amasyali & El-Gohary, 2018). In recent years, the data-driven approach has been gaining attention because of its high efficiency and accuracy. One of the biggest advantages of the data-driven model is that there is no need to build complex physical models. The relationships between input and output variables are captured automatically through advanced data analytics such as machine learning and artificial intelligence. In this sense, the data-driven model is more efficient and flexible. A data-driven model comprises three major parts: input features, training algorithms, and output features. The input features for a data-driven model should incorporate all major driving factors of the output variables. Insufficient input features will decrease model precision, whereas too many redundant input variables may cause model overfitting and increase execution time (Zhao & Magoulès, 2012). Thus, model input selection is vital for building an efficient data-driven model. As building load is closely related outdoor weather condition and usage, variables such as dry-bulb temperature, wet-bulb temperature, relative humidity, wind speed, solar radiation and occupancy schedule were typically selected as model inputs (Benedetti, Cesarotti, Introna, & Serranti, 2016; Yang, Rivard, & Zmeureanu, 2005) Input features of previous studies are summarized in Table 1. Besides from directly using measured data as input variables, several studies transformed the raw data into a more

* Corresponding author.
E-mail address: xupeng@tongji.edu.cn (P. Xu).

https://doi.org/10.1016/j.scs.2019.101698
Received 16 February 2019; Received in revised form 25 June 2019; Accepted 3 July 2019
Available online 18 July 2019

Table 1
Summary of data-driven models for building energy prediction.

Reference	Input features	Output features	Training Algorithm
Li, Ding, Lǔ, Xu, and Li (2010)	Date, daily average temperature, daily lowest temperature, daily highest temperature	Hourly cooling energy	Support vector machine (SVM), Principal component analysis SVM (PCA-SVM), Kernel PCA-SVM (KPCA-SVM)
Li, Meng, Cai, Yoshino, and Mochida (2009)	Dry-bulb temperature, relative humidity, solar radiation	Hourly cooling energy	SVM, Artificial neural network (ANN)
Li, Lǔ, Ding, Xu, and Li (2009)	Dry-bulb temperature, relative humidity, solar radiation	Hourly cooling energy	Least squares SVM (LS-SVM), ANN
Yun, Luck, Mago, and Cho (2012)	Historical cooling energy	Hourly cooling/heating energy	Average regression (AR)
Ahmad and Chen (2019)	Wet bulb temperature, dew point temperature, pressure, relative humidity	Hourly cooling and heating energy	nonlinear autoregressive model(NARM), linear model using stepwise regression (LMSR), random forest(RF)
Chou and Bui (2014)	Relative compactness, surface area, wall area, roof area, overall height, orientation, glazing area, glazing area distribution	Hourly cooling/heating energy	ANN, SVM, Decision tree (DT), etc.
Wang, Lu, and Li (2019)	Dry bulb temperature, dew point temperature, relative humidity, wind speed, solar radiation, time of day	Hourly cooling and heating energy	extreme gradient boosting (XGBoost), RF, ANN) gradient boosting decision tree (GBDT), SVR
Safa, Safa, Allen, Shahi, and Haas (2017)	Average monthly temperature, full-time employee	Monthly cooling and heating energy	MLR, ANN
Penya, Borges, and Fernández (2011)	Day of week, type of day, season, wind direction, humidity, perception, sigma direction, sigma speed, air temperature, average speed	Hourly cooling and heating energy	AR, ANN, SVM, Bayesian network
Liu and Chen (2013)	Number of people in building, solar radiation	Lighting energy	SVM, ANN

compact format as model input. This process is called feature transformation (Bourdeau, Zhai, Nefzaoui, Guo, & Chatellier, 2019). The feature transformation methods can be classified into two types, i.e., engineering and statistical. Engineering feature extraction is based on engineering experience. Grolinger, L'Heureux, Capretz, and Seewald (2016) used the data at previous time step as model input considering building thermal capacity. Cui, Wu, Hu, Weir, and Li, (2016) used both specification data and lagged data as model input features and pointed out that the combined features gave superior performance over models which built solely on lagged data. Statistical method uses statistical methods to reduce feature dimension. Summarizing statistical indicators such as maximum, minimum or medium values of the raw data are often used as model input features (Lemke & Gabrys, 2010). This method is easy to interpret and operate. Another commonly used statistical method is principle component analysis (PCA). PCA uses an orthogonal transformation to convert a set potentially correlated variables into a set of unrelated variables called principal components. The principle components have no physical mean but contain major information of the original data (Jolliffe, 2002). PCA is applicable for circumstances when feature space is large and it is hard to select dominant features manually. Li et al. (2018) applied PCA for relevant input feature selection and then used a combined model for building electrical energy consumption prediction. Nilashi et al. (2017) implemented PCA with an adaptive network-based fuzzy inference system for residential building load prediction.

A machine learning algorithm is the core of a prediction model. Algorithms such as multi-variable linear regression (MLR), support vector machine (SVM), artificial neural network (ANN), decision tree (DT), random forest (RF) and extreme gradient boosting (XGBoost) are commonly used as training algorithms. The first four of aforementioned algorithms belongs to single prediction algorithms. The last two are ensemble algorithms. Several studies suggest that ensemble algorithms are more effective than single prediction algorithms (Foucquier, Robert, Suard, Stéphan, & Jay, 2013; Martínez-álvarez, Troncoso, Asencio-Cortés, & Riquelme, 2015). Table 1 summarizes machine learning algorithms used in previous studies.

In previous studies, a number of meteorological parameters were adopted as model input features. However, parameters such as relative humidity, solar radiation, and wind speed are often excluded from weather forecast reports. Thus, it is impractical to use such complicated models for energy prediction. This paper proposes a simplified prediction model that can be operated easily and yield accurate results. Weather data is transformed into degree-day as model input features. Different from traditional degree-day methods based on fixed balance point, the degree-day used in this study is based on energy profile pattern of a specific building. Compared to the structure of previous models, the structure of this model is simpler and more concise. This paper adopted four machine learning algorithms, i.e., MLR, SVM, and ANN owing to their wide application and high performance. The prediction accuracy of four algorithms is discussed in Section 4. The output variable is the daily electricity consumption (kWh) of an HVAC system. This approach is applied to predict the energy consumption of an HVAC system of a large commercial building. The prediction performance indicates that this approach is suitable for use in engineering applications.

2. Research methodology

2.1. Selection of input features

Building energy is mainly influenced by five factors: (1) the climate and location of the building, (2) the desired temperature and humidity, (3) the number of occupants and period of occupancy, (4) the thermal performance of the structure itself, and (5) the building use (Zhang, Cao, & Romagnoli, 2018). For this research, energy prediction is conducted on a specific building and no retrofitting work is carried out during the analysis period. Thus, factors 1 and 4 are excluded from this prediction model. Factors 2, 3, and 5 are the driving forces of the energy variance in a specific building. In previous studies, meteorological parameters such as dry-bulb temperature, relative humidity ratio, wind speed, and solar radiation were used as prediction indicators. However, most of these parameters are difficult to obtain. To reduce the computation burden and avoid overfitting, we will not take all meteorological parameters into consideration. A correlation analysis between HVAC energy and weather parameters is performed first, to determine the most prominent parameters. In this research, we adopt the Pearson correlation coefficient as the indicator to represent how closely two variables are related. For a given pair of random variables (X, Y), the Pearson correlation coefficient can be easily calculated using Eqs. (1–4).

$$R = \sigma_{xy}^2 / \sqrt{\sigma_x^2 \sigma_y^2} \tag{1}$$

$$\sigma_{xy}^2 = \frac{1}{n}\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y}) \tag{2}$$

Table 2
Correlation coefficients (R) of HVAC system electricity consumption and meteorological parameters.

	Dry-bulb temperature	Relative humidity ratio	Wind speed	Solar radiation
Summer	0.91	−0.21	−0.05	0.47
Winter	−0.83	−0.15	0.21	0.07

$$\sigma_x^2 = \frac{1}{n}\sum\nolimits_{i=1}^{n}(x_i - \bar{x})^2 \tag{3}$$

$$\sigma_y^2 = \frac{1}{n}\sum\nolimits_{i=1}^{n}(y_i - \bar{y})^2 \tag{4}$$

where R is the Pearson correlation coefficient, σ_{xy} is the covariance of X and Y, σ_x is the standard deviation of X, σ_y is the standard deviation of Y, $\bar{x}$ is the mean of X, $\bar{y}$ is the mean of Y, and x_i and y_i are the individual sample points indexed with i.

The Pearson correlation coefficients between the electricity consumption of an HVAC system and four meteorological parameters are calculated, and listed in Table 2. Data of the electricity consumed by an HVAC system and the four meteorological parameters used here are the same as those in the case study discussed in Section 3. A positive value indicates that the two variables are positively correlated, and vice versa. A higher value indicates a stronger relationship. It is evident that dry-bulb temperature is the factor that is most relevant to the energy variance of an HVAC system, in both summer and winter. The absolute value of the correlation coefficients between the other three parameters and the energy consumption of an HVAC system is less than 0.5, so they are excluded from the prediction model.

Another factor that influences energy consumption is the building internal load, which is largely determined by the building usage pattern of density and occupancy schedules, lighting, etc. However, it is difficult to obtain data such as occupancy schedules. For commercial buildings such as office and retail buildings, the usage pattern is regular. In other words, the usage pattern for a certain day is similar to that for all similar days. For example, the usage pattern on a Sunday will be similar to that on all Sundays. For simplicity, the day type and month type are used to represent the building usage pattern. The research outline of this model is shown in Fig. 1.

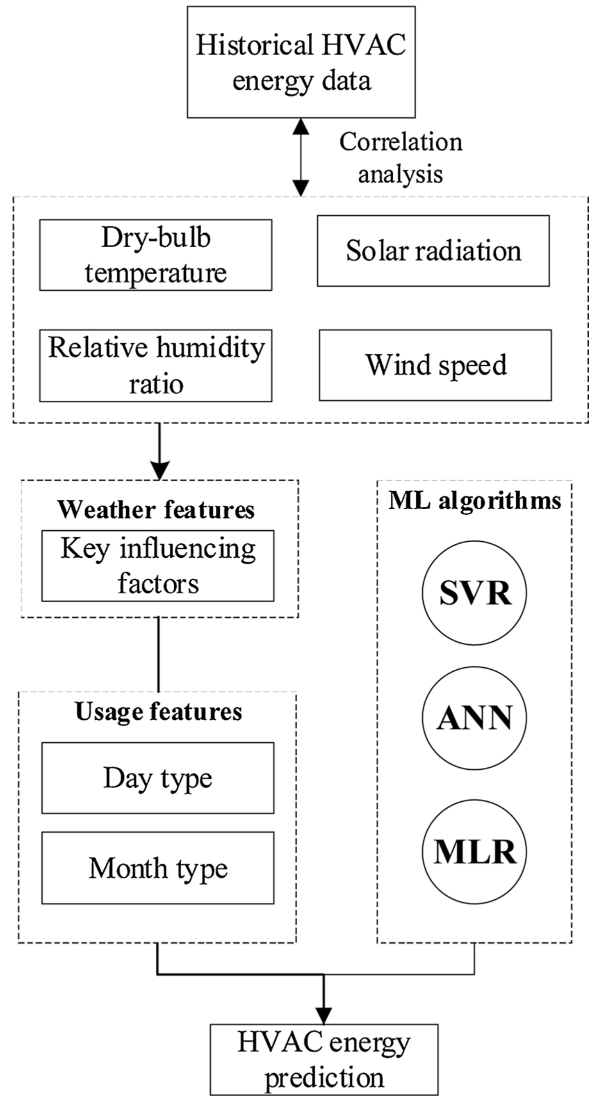

Fig. 1. Research outline of energy prediction model.

2.2. Prediction algorithms

In this study, three popular machine learning algorithms are selected: SVM, ANN, and MLR. The first two are capable of capturing complex nonlinear relationships between input and output variables. MLR is one the most commonly used methods for linear regression analysis. It is selected as a performance benchmark.

This paper uses the extension algorithm of SVM, i.e. support vector regression(SVR). SVM is a supervised machine learning algorithm that is used as a classification method. It aims to find a hyperplane that separates samples of different labels with maximal margins in the original space or a higher dimensional space, using a kernel function for mapping. Commonly used kernel functions include linear functions, polynomial functions, and Gaussian functions. The SVR employs the same principles as the SVM, with a few minor differences. In SVR, a loss function is defined that ignores errors situated within a certain distance of the true value. This type of function is often called an epsilon-insensitive loss function (Smola & Schölkopf, 2004).

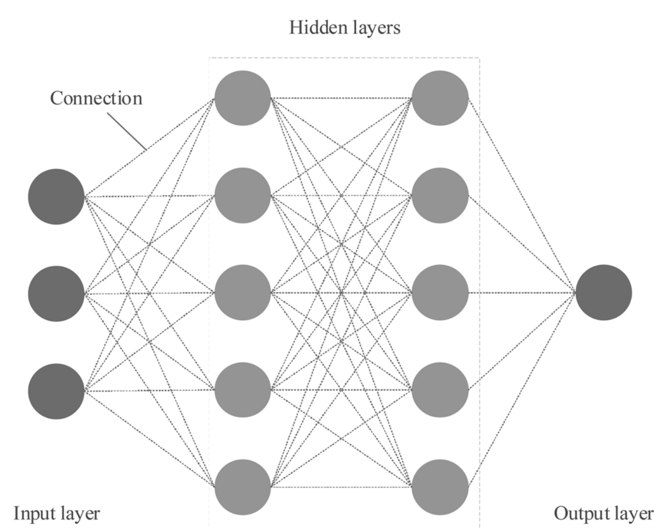

Fig. 2. Typical structure of ANN.

ANN is also a popular algorithm that can solve both linear and nonlinear problems. It is usually composed of a number of connected units or nodes organized in several layers. A typical neural network structure is shown in Fig. 2. Once the structure is determined, the weights and parameters can be trained through a back propagation (BP) algorithm. Thus, it is sometimes called a BP network. ANN has been proven useful in solving problems of pattern recognition, prediction, and function approximation (Ruppert, 2004).

MLR is an extension of simple linear regression analysis. It attempts to model the relationship between several independent variables and a single dependent variable by fitting a linear equation to the observed data. MLR is usually given in the following form:

$$\mu = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p \tag{5}$$

where μ is the dependent variable to be predicted, $x_1, x_2 \ldots x_p$ are p independent explanatory variables, $b_1, b_2 \ldots b_p$ are the estimated regression coefficients, and b_0 is the intercept. In addition to capturing the relationship between explanatory and prediction variables, MLR can determine the explanatory variables that are more important than others. MLR can be easily implemented and interpreted; hence, it is very widely used in research. However, MLR does not perform well when the explanatory and prediction variables are non-linearly correlated.

3. Case study

3.1. Data description

The energy data to be analyzed in this study are obtained from a large retail building in Nanjing, China. The data are a daily time series collected from 1/1/2016 to 8/30/2018. The data includes the total electricity consumption of various units including chillers, chilled water pumps, cooling water pumps, cooling towers, and heat pumps. The historical energy data profile is shown in Fig. 3. As there is no energy consumed by the HVAC system during the transition season, this period is not considered. In Nanjing, the transition season is in April and November. The cooling season is from May to October. The rest of the months are considered part of the heating season. The meteorological data are obtained from the National Meteorological Information Center (Anonymous, 2019).

3.2. Development of energy prediction model

As described in Section 2.2, a correlation analysis is conducted first to derive the meteorological parameters most relevant to the energy consumed by an HVAC system. The dry-bulb temperature is selected as one of the input features for predicting the energy consumed by an HVAC system. However, we will not directly use the dry-bulb temperature as the model input feature. Instead, the dry-bulb temperature is transformed into degree-day for analysis. Degree-day method is a simple but effective method for building energy analysis (ASHRAE, 2009). Degree-daycompares the mean (the average of the high and low) outdoor temperatures recorded at a location with a balance point temperature. The balance point temperature is defined as that value of the outdoor temperature when total heat loss is equal to heat gain of a building. A high number of degree-days generally results in higher levels of energy use for space heating or cooling. Balance point temperature is influenced by building thermal characteristics and usage pattern. Thus, the balance point temperature may to different building to building. It is also hard to calculate mathematically. Here we proposed an intuitive method to obtain the balance point temperature. When we plot the daily electricity consumption data of an HVAC system against the daily mean temperature (i.e., mean value of the maximum and minimum temperature), we find that when the mean temperature is higher than 15 °C or lower than 10 °C, the electricity consumption of the HVAC system varies almost linearly with the temperature variation (Fig. 4). Thus, $T_c = 15$ °C and $T_h = 10$ °C are selected as the cooling and heating standard temperatures, respectively. Therefore, the daily cooling degree-day (CDD) and heating degree-day (HDD) are calculated using Eqs. (5) and (6), as energy prediction model input features:

$$\text{CDD} = \max\left((T_{max} + T_{min})/2 - T_c\right),\ 0) \tag{6}$$

$$\text{HDD} = \max\left(T_h - (T_{max} + T_{min})/2\right),\ 0) \tag{7}$$

where T_{max} and T_{min} are the maximum and minimum hourly temperature of a day, respectively.

As discussed in Section 2.2, the day type and month type are also used as indicators for energy prediction. Integers 1–8 are adopted to represent Monday to Sunday and holidays, whereas 1–12 are used to represent January to December. Accordingly, for this research, the energy prediction model includes three input features: CDD (for summer) or HDD (for winter), day type (1–8), and month type (1–12). The model output is the daily electricity consumption of the HVAC system. Compared to the models used in previous studies, this model is largely simplified.

The entire dataset is divided into training and testing data in proportions of 70% and 30%, respectively. The model hyperparameters are found through minimizing five-fold cross-validation loss. For SVR, the parameters to be optimized are the complexity parameter C, smoothing parameter sigma, and kernel function. C controls the structure complexity of the model. A larger C would increase the likelihood of the model overfitting, and vice versa. The parameter sigma controls the model decision boundary. A larger sigma leads to a more flexible and smoother boundary. Three kernel functions (i.e., Gaussian, linear, and polynomial) form the decision pool, from which the most suitable decision is chosen at each calculation step. For ANN, the model structure, as determined by the number of hidden layers and neurons of each layer, has a prominent influence on the model performance. In this

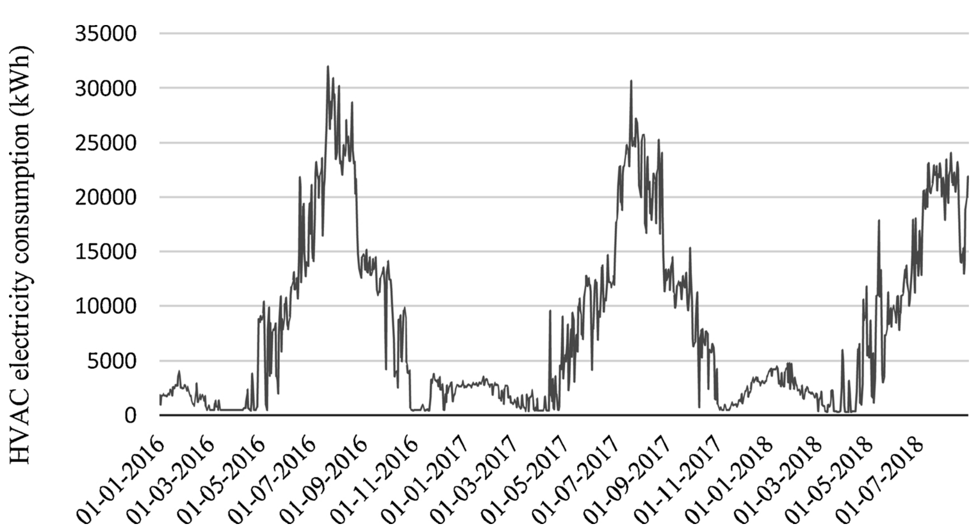

Fig. 3. Profile of historical HVAC system electricity consumption.

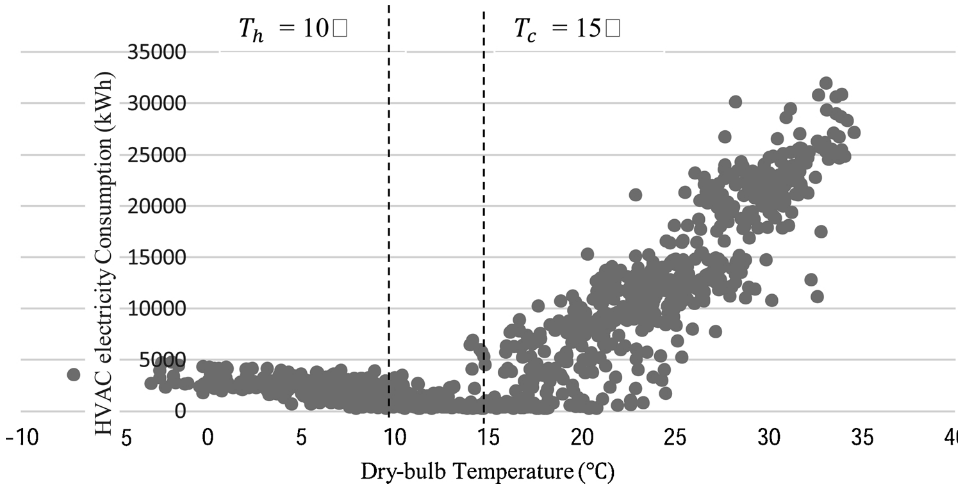

Fig. 4. Pattern of HVAC system energy consumption.

Table 3
Training and testing performance of three algorithms (using degree-day as input feature).

Model Type		MAE		RMSE		CV-RMSE	
		Cooling	Heating	Cooling	Heating	Cooling	Heating
SVR	Training	1757	377	2399	539	17.9%	22%
	Testing	2542	644	3137	855	19.3%	32.3%
ANN	Training	1819	535	2352	668	17.6%	27.5%
	Testing	2672	736	3327	889	20.6%	33.6%
MLR	Training	2337	536	3074	656	23%	20%
	Testing	2552	709	3341	891	21%	31.4%

research, we obtain the optimized model structure (i.e., two hidden layers and four neurons in each layer) through manual adjustment.

4. Results and discussions

The model prediction performance is evaluated using three indexes, as defined in Eqs. (7–9). They are the root mean squared error (RMSE), mean absolute error (MAE), and coefficient of variation of the root mean squared error (CV-RMSE). The first two indexes are scale-dependent, whereas CV-RMSE is scale-independent. Scale-independent metrics can be used to evaluating performance with other studies. Previous studies have specified that model of CV-RMSE below 30% is sufficient for engineering purpose (Reddy, Maor, & Panjapornpon, 2007). For a given variable noted as Y and the predicted variable noted as $\hat{Y}$:

$$\text{MAE}=\frac{\sum_{k=1}^{n}|y_k-\hat{y}_k|}{n} \tag{8}$$

$$\text{RMSE}=\sqrt{\frac{\sum_{k=1}^{n}(y_k-\hat{y}_k)^2}{n}} \tag{9}$$

$$\text{CV}-\text{RMSE}=\sqrt{\frac{\sum_{k=1}^{n}(y_k-\hat{y}_k)^2}{n}}\bigg/\frac{\sum_{k=1}^{n}y_k}{n} \tag{10}$$

where y_k is the value of the k th point of Y, $\hat{y}_k$ is the value of the k th point of $\hat{Y}$, and n is the sample number of Y.

Tables 2 and 3 summarize the training and testing performance in terms of the obtained MAE, RMSE, and CV-RMSE. Figs. 5 and 6 intuitively show the model performance during the cooling and heating seasons. The training and test results are plotted in the same figure for convenience, using a dotted line as a boundary.

Table 2 shows that the prediction accuracy of a model based on average dry-bulb temperature is lower than that based on degree day. This proves that degree day is more suitable as an indicator for predicting energy consumption. From Table 3, it can be seen that SVR and ANN perform better than MLR in predicting the energy consumed by an HVAC system. This is because a building thermal process is a complex nonlinear process. MLR is too simple an algorithm to map a nonlinear relationship between input and output variables. The SVR prediction

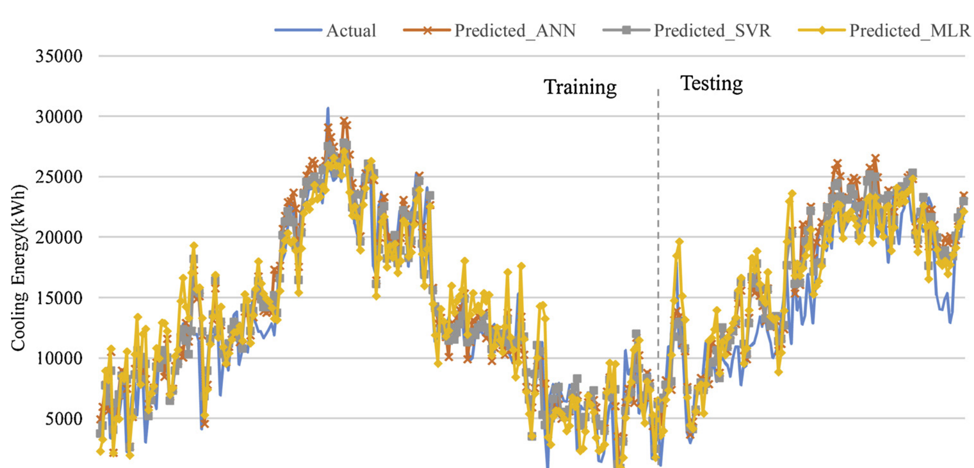

Fig. 5. Prediction performance during cooling season.

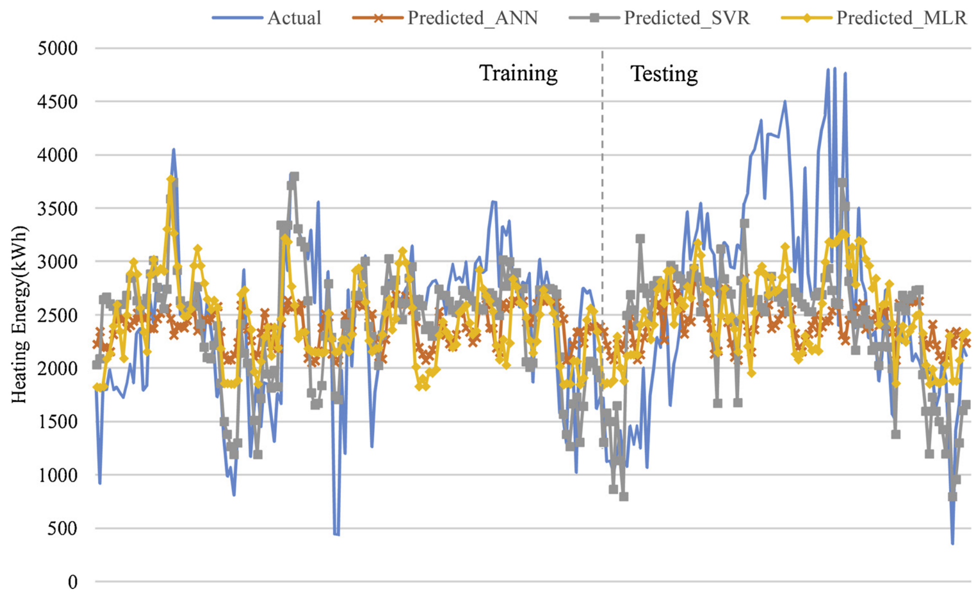

Fig. 6. Prediction performance during heating season.

performance is slightly better than that of ANN because SVR is more suitable for cases with a small dataset. However, the discrepancy between the SVR and ANN cooling models is very small, and hence, it can be ignored in engineering applications. In addition, the cooling CV-RMSE values of SVR are less than 20%, which means that the model is sufficiently suitable for engineering applications (ASHRAE, 2002; FEMP, 2008).

It is evident that for all three algorithms, the cooling model performs much better than the heating model. The main reason for this is that the heating data are insufficient. For this study, 505 samples were collected during the summer season, and only 261 samples during the heating season. The model prediction performance degrades rapidly as the amount of data decreases. Another interesting phenomenon is that MLR performs better than SVR and ANN during the heating season. The main reason for this is that the SVR and ANN models are more likely to overfit than the MLR model when the data is insufficient. Thus, sufficient data is vital for building a reliable data-driven model.

5. Conclusions

Predicting the energy consumed by a building is essential for improving its energy efficiency in terms of building operation optimization and fault detection and diagnosis. This paper proposed a simplified energy prediction method for engineering applications. For this method, only three variables were adopted as model input features. The first one was weather related data. Instead of directly using average daily dry-bulb temperature as the model input feature, we innovatively transformed it to degree-day. The result showed that using degree-day as input had better performance. It should be noted that the balance point temperature for calculating degree-day should not be a fixed value, because it is influence by building characteristics and usage pattern. This paper also proposed an intuitive way to obtain the balance point temperature. Another two features were the building usage characteristics represented by day type and month type. Although the model structure and input features used in this study were largely simplified as compared to the models employed in previous studies, its prediction accuracy is favorable. This study adopted three popular machine learning algorithms (i.e., MLR, SVR, and ANN) as prediction models. The results showed that the SVR and ANN models performed better than the MLR model, indicating that a building thermal process is usually nonlinear and complex. When trained with sufficient data, the CV-RMSE of the SVR model was less than 20%, proving that this approach can be applied in practical applications. The large discrepancy between the cooling and heating energy prediction performance indicated that the size of the training dataset is vital for model prediction performance. All three methods exhibited poor performance in heating energy prediction; however, their performance can be improved if the training dataset is large.

Acknowledgment

This research is fully supported by the Shanghai Zhangjiang National Independent Innovation Demonstration Zone Special Development Fund Project, 201705-MH-CHJ-C1086-009 and Shanghai Minhang District Science and Technology Committee Fund Project, 2017MH314.

References

Ahmad, T., & Chen, H. (2019). Nonlinear autoregressive and random forest approaches to forecasting electricity load for utility energy management systems. *Sustainable Cities and Society, 45*, 460–473. https://doi.org/10.1016/j.scs.2018.12.013.

Amasyali, K., & El-Gohary, N. M. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews, 45*, 460–473. https://doi.org/10.1016/j.rser.2017.04.095.

http://data.cma.cn/site/index.html.

ASHRAE (2002). *ASHRAE guideline 14: Measurement of energy and demand savings.* American Society of Heating, Ventilating, and Air Conditioning.

ASHRAE (2009). *ASHRAE handbook—Fundamentals.* Atlanta, GA, USA: Refrigerating American Society of Heating and Air-Conditioning Engineers.

Benedetti, M., Cesarotti, V., Introna, V., & Serranti, J. (2016). Energy consumption control automation using Artificial Neural Networks and adaptive algorithms: Proposal of a new methodology and case study. *Applied Energy, 165*, 60–71. https://doi.org/10.1016/j.apenergy.2015.12.066.

Bourdeau, M., Zhai, X., Nefzaoui, E., Guo, X., & Chatellier, P. (2019). Modeling and forecasting building energy consumption: A review of data-driven techniques. *Sustainable Cities and Society, 48*, 1015–1033. https://doi.org/10.1016/j.scs.2019.101533.

Chou, J. S., & Bui, D. K. (2014). Modeling heating and cooling loads by artificial intelligence for energy-efficient building design. *Energy and Buildings, 82*, 437–446. https://doi.org/10.1016/j.enbuild.2014.07.036.

Cui, C., Wu, T., Hu, M., Weir, J. D., & Li, X. (2016). Short-term building energy model recommendation system: A meta-learning approach. *Applied Energy, 172*, 251–263. https://doi.org/10.1016/j.apenergy.2016.03.112.

FEMP (2008). *M & V guidelines: Measurement and verification for federal energy projects. Version 3.0*.

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. https://doi.org/10.1016/j.rser.2013.03.004.

Grolinger, K., L'Heureux, A., Capretz, M. A. M., & Seewald, L. (2016). Energy forecasting for event venues: Big data and prediction accuracy. *Energy and Buildings, 112*,

222–233. https://doi.org/10.1016/j.enbuild.2015.12.010.
Jolliffe, I. T. (2002). *Principal component analysis* (2nd ed.). Springer, Ed.
Lemke, C., & Gabrys, B. (2010). Meta-learning for time series forecasting and forecast combination. *Neurocomputing. 73*, 2006–2016. https://doi.org/10.1016/j.neucom.2009.09.020.
Li, K., Xie, X., Xue, W., Dai, X., Chen, X., & Yang, X. (2018). A hybrid teaching-learning artificial neural network for building electrical energy consumption prediction. *Energy and Buildings, 174*, 323–334. https://doi.org/10.1016/j.enbuild.2018.06.017.
Li, X., Bowers, C. P., & Schnier, T. (2010). Classification of energy consumption in buildings with outlier detection. *IEEE Transactions on Industrial Electronics, 57*, 3639–3644. https://doi.org/10.1109/TIE.2009.2027926.
Li, X., Ding, L., Lǔ, J., Xu, G., & Li, J. (2010). A novel hybrid approach of KPCA and SVM for building cooling load prediction. *In 3rd International Conference on Knowledge Discovery and Data Mining, WKDD 2010*. https://doi.org/10.1109/WKDD.2010.137.
Li, X., Lǔ, J. H., Ding, L., Xu, G., & Li, J. (2009). Building cooling load forecasting model based on LS-SVM. *In Proceedings - 2009 Asia-Pacific Conference on Information Processing, APCIP 2009*. https://doi.org/10.1109/APCIP.2009.22.
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*, 2249–2256. https://doi.org/10.1016/j.apenergy.2008.11.035.
Liu, D., & Chen, Q. (2013). Prediction of building lighting energy consumption based on support vector regression. *2013 9th Asian Control Conference, ASCC 2013*. https://doi.org/10.1109/ASCC.2013.6606376.
Martínez-álvarez, F., Troncoso, A., Asencio-Cortés, G., & Riquelme, J. C. (2015). A survey on data mining techniques applied to electricity-related time series forecasting. *Energies. 8*, 13162–13193. https://doi.org/10.3390/en81112352.
Nilashi, M., Dalvi-Esfahani, M., Ibrahim, O., Bagherifard, K., Mardani, A., & Zakuan, N. (2017). A soft computing method for the prediction of energy performance of residential buildings. *Measurement: Journal of the International Measurement Confederation, 109*, 268–280. https://doi.org/10.1016/j.measurement.2017.05.048.
Penya, Y. K., Borges, C. E., & Fernández, I. (2011). Short-term load forecasting in non-residential buildings. *In IEEE AFRICON Conference*. https://doi.org/10.1109/AFRCON.2011.6072062.
Pérez-Lombard, L., Ortiz, J., & Pout, C. (2008). A review on buildings energy consumption information. *Energy and Buildings, 40*, 392–398. https://doi.org/10.1016/j.enbuild.2007.03.007.
Reddy, T., Maor, I., & Panjapornpon, C. (2007). Calibrating detailed building energy simulation programs with measured data— Part 2: Application to Three case Study \nOffice buildings. *American Society of Heating, Refrigerating and Air-Conditioning Engineers HVAC & R Research, 13*, 243–265.
Ruppert, D. (2004). The elements of statistical learning: Data mining, inference, and prediction. *Journal of the American Statistical Association, 99*, 466–567. https://doi.org/10.1198/jasa.2004.s339.
Safa, M., Safa, M., Allen, J., Shahi, A., & Haas, C. T. (2017). Improving sustainable office building operation by using historical data and linear models to predict energy usage. *Sustainable Cities and Society, 29*, 107–117. https://doi.org/10.1016/j.scs.2016.12.001.
Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing, 14*, 199–222. https://doi.org/10.1023/B:STCO.0000035301.49549.88.
Ürge-Vorsatz, D., Cabeza, L. F., Serrano, S., Barreneche, C., & Petrichenko, K. (2015). Heating and cooling energy trends and drivers in buildings. *Renewable and Sustainable Energy Reviews, 41*, 85–98. https://doi.org/10.1016/j.rser.2014.08.039.
Wang, R., Lu, S., & Li, Q. (2019). Multi-criteria comprehensive study on predictive algorithm of hourly heating energy consumption for residential buildings. *Sustainable Cities and Society, 39*, 1016–1023. https://doi.org/10.1016/j.scs.2019.101623.
Yang, J., Rivard, H., & Zmeureanu, R. (2005). On-line building energy prediction using adaptive artificial neural networks. *Energy and Buildings, 37*, 1250–1259. https://doi.org/10.1016/j.enbuild.2005.02.005.
Yun, K., Luck, R., Mago, P. J., & Cho, H. (2012). Building hourly thermal load prediction using an indexed ARX model. *Energy and Buildings, 54*, 225–233. https://doi.org/10.1016/j.enbuild.2012.08.007.
Zhang, C., Cao, L., & Romagnoli, A. (2018). On the feature engineering of building energy data mining. *Sustainable Cities and Society, 39*, 508–518. https://doi.org/10.1016/j.scs.2018.02.016.
Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. *Renewable and Sustainable Energy Reviews, 16*, 3586–3592. https://doi.org/10.1016/j.rser.2012.02.049.