Applied Energy 289 (2021) 116652

Contents lists available at ScienceDirect

Applied Energy

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

Development of a multi-granularity energy forecasting toolkit for demand response baseline calculation

Huajing Sha [a], Peng Xu [a,*], Meishun Lin [b], Chen Peng [b], Qiang Dou [b]

[a] Department of Mechanical and Energy Engineering, Tongji University, Shanghai 201804, China
[b] Persagy Technology Co., Ltd, 100096 Beijing, China

HIGHLIGHTS

- Review of feature engineering research for HVAC energy forecasting models.
- A novel feature engineering method for exploring informative features.
- An easy-to-use, high-accuracy toolkit for demand response baseline calculation.
- Comparative tests verify the stability and accuracy of this energy prediction toolkit.
- The average CV-RMSE of the target models for hourly energy prediction is <8%.

ARTICLE INFO

Keywords:
HVAC energy forecasting
Demand response
Baseline load
Machine learning
Feature engineering

ABSTRACT

The peak load caused by heating, ventilation, and air-conditioning (HVAC) systems is one of the main control targets of a demand response (DR) program. One key issue related to DR is the baseline energy consumption forecasting based on which the DR strategies and performance can be evaluated. Data-driven models, as a promising method for HVAC energy prediction, have been widely studied. But most existing researches have focused on developing complicated algorithms rather than exploring informative features. In this study, a comprehensive review of feature engineering for HVAC energy prediction model development is presented. A novel feature engineering method is roposed. Besides, an easy-to-use, high-accuracy HVAC energy forecasting toolkit that is applicable to datasets of various granularities is developed. This toolkit uses easily available meteorological parameters and raw historical energy data as inputs, on which it performs data preprocessing, feature extension, and integrated optimization, thereby producing the predicted data. By employing a novel feature extension strategy and integrated optimization of feature selection and hyperparameter tuning, this toolkit performs capably in terms of prediction accuracy and stability. The results of a comparative experiment conducted on large-scale data verify that the average forecasting error (measured in terms of the coefficient of variation of the root mean square error) is <8%.

1. Introduction

Demand response (DR) refers to incentives or programs designed to motivate end-users to adapt their normal electricity use behavior when grid stability is jeopardized. By shifting energy packages from on-peak periods to other periods, DR effectively ensures grid reliability and reduces grid capacity. DR is currently garnering increasing attention in China with the escalating electricity demand. Several active and passive DR strategies have been proven to be feasible in curtailing peak loads. However, a key challenge to be addressed for DR implementation is baseline estimation, which enables DR performance to be measured. DR performance is calculated as the reduction in electricity consumption compared with the baseline load, which is an estimate of electricity that would be consumed by end-users in the absence of demand curtailment strategies. Accurate calculation of the baseline load not only supports fair compensation of DR participants but also provides useful information to system operators. In commercial buildings, where the peak load is mainly caused by the heating, ventilation, and air-conditioning (HVAC) electricity demand, it is important to predict the future HVAC load accurately.

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

https://doi.org/10.1016/j.apenergy.2021.116652
Received 30 November 2020; Received in revised form 7 February 2021; Accepted 11 February 2021
Available online 2 March 2021

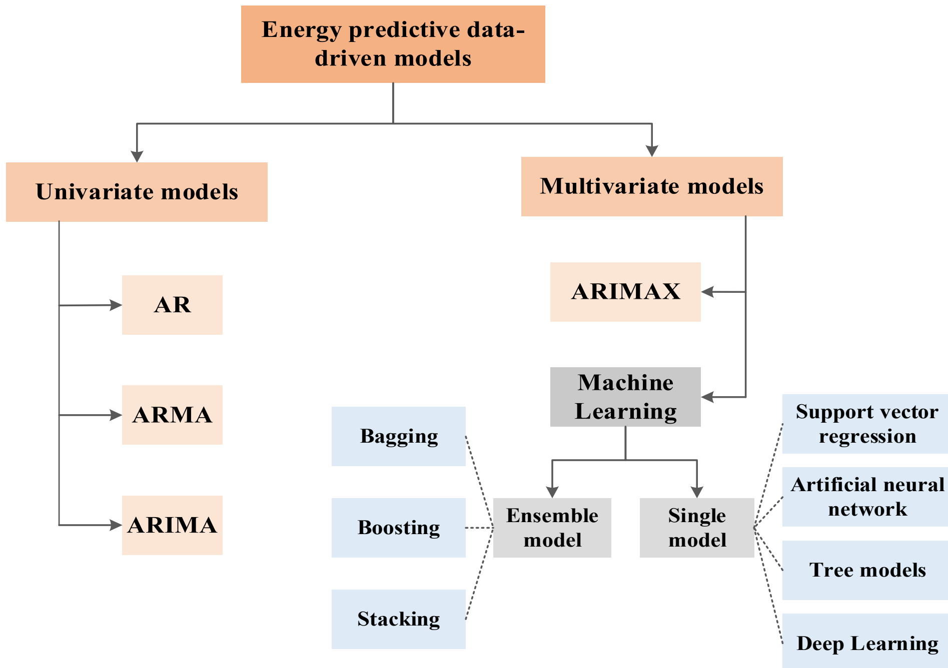

Fig. 1. Classification of data-driven models for building energy prediction.

Two main types of HVAC energy forecasting methods exist depending on whether physical or data-driven models are used. Data-driven models are superior to physical models in terms of convenience and prediction accuracy. Developed with large amounts of historical operational data to capture the complex and nonlinear relationship between input and output variables, data-driven models are devoid of detailed geometrical information regarding buildings and often provide more accurate results. Feature engineering, the most important step for data-driven model development, refers to the process of creating features that enable algorithms to function through the use of domain knowledge. Feature engineering determines the upper limit of model performance. Well-engineered features can help achieve highly accurate predictions with simple algorithms [1]. However, most existing studies have focused on developing complicated algorithms rather than on feature engineering. The features employed in the most studies are limited to directly observable meteorological parameters and time indices.

In this study, we present a comprehensive review of existing energy prediction research for DR program and feature engineering research on data-driven models for HVAC energy forecasting. We then propose an HVAC energy forecasting toolkit. This toolkit requires easily available data as inputs but offers a higher prediction accuracy owing to its novel feature engineering strategy. It is easy to use for both DR participants and system operators, and it can also be used for other practices involving HVAC load forecasting, such as district energy planning [2], building retrofitting [3], fault detection and diagnosis for energy systems [4], and energy policy making [5]. As this toolkit is free of using building characteristics, it is applicable for residential sector as well as commercial sector. With the rapid development of Internet of Things (IoT), more smart devices will be embedded into buildings. It is encouraged to integrate this toolkit with IoT system for DR program implementation in both smart homes and commercial buildings. The data collected from sensors and meters deployed in IoT systems can be sent to this energy forecasting toolkit for knowing energy demand in advance [6].

This paper is organized as following structure: Section 2 presents a detailed review on energy prediction models and feature engineering strategies employed by previous studies. A detailed description of the toolkit development is provided in Section 3. A comparative experiment was conducted using data collected from 20 commercial buildings to validate the usefulness of the toolkit. The data and comparative baseline models are described in Section 4. The calculation results are discussed in Section 5. In Section 6, we present our conclusions. The contributions and novelty of this study can be summarized as follows:

The majority of existing research on energy prediction models focuses on new and complicated machine learning algorithms. This study emphasizes and proves that feature engineering rather than algorithm plays a decisive role in performance of data-driven energy prediction models. A comprehensive review of features used in existing energy prediction models are analyzed and summarized.

A novel feature engineering method along with integrated feature selection and model hyperparameter optimization mechanism are proposed based on which a compact toolkit for DR baseline calculation is developed. This toolkit is easy to use and is suitable for a variety of application scenarios.

The accuracy and stability of the proposed DR baseline calculation toolkit are validated to be fairly acceptable based on large-scale commercial building dataset.

2. Literature review

2.1. Existing energy forecasting research for DR program

DR can be taken as a set of methods that help balancing supply and demand. Energy prediction for supply side and demand side are equally important especially as unstable renewable energy sources are increasingly applied by DR programs to reduce peak demand. On the one hand, knowing demand load in advance can help to develop suitable strategies for end-users. On the other hand, predicting the amount of energy to be supplied during the upcoming period can help to contribute a better management of disparity between supply and demand [6]. Researchers have tried various novel methods and algorithms to improve predicting accuracy for both scenarios. Kebir et al. employed backpropagation combined to chi-squared method for weighting historical data to predict short-term demand peak load [7]. Alduailij et al. compared the prediction accuracy of different data-driven models including linear regression, dynamic regression, ARIMA, artificial neural network, and deep neural network, and found that ARIMA with exogenous variables outperformed all other models [8]. In the project of demand response in

Table 1
Summary of input features for HVAC energy forecasting models.

Ref	Input features							Output	Task type
	Meteorological parameters	Time index	Occupancy	Building or system characteristics	Lagged Meteorological parameters	Lagged load/ energy	Others		
[51]	Y	Y	N	Y	N	N	/	Yearly energy use	Parallel
[52]	Y	Y	N	Y	N	Y	/	The 24-h ahead power consumption	Parallel
[53]	Y	Y	N	N	N	N	/	Hourly energy consumption	Sequential
[54]	Y	Y	N	N	N	Y	/	The 24-h ahead power consumption	Sequential
[37]	Y	Y	N	N	Y	Y	/	Hourly energy consumption	Sequential
[40]	Y	N	N	Y	N	N	/	Energy prediction after two hours	Sequential
[55]	Y	Y	N	N	N	Y	/	Hourly energy consumption	Sequential
[27]	Y	Y	N	N	Y	Y	/	Half-hourly energy consumption	Sequential
[56]	N	N	N	Y	N	N	/	Heating/cooling load	Parallel
[57]	Y	Y	Y	N	N	N	System operation parameters	Hourly energy power	Sequential
[58]	Y	Y	N	N	N	N	System operation parameters	Hourly energy consumption	Sequential
[44]	Y	N	N	Y	N	N	/	Weekly heating energy consumption	Sequential
[59]	Y	Y	N	N	N	N	/	Hourly energy consumption	Sequential
[41]	Y	Y	N	N	N	N	Indoor environment parameters	10-min energy consumption	Sequential
[60]	N	Y	N	N	N	Y	/	Hourly energy consumption	Sequential
[32]	Y	N	N	N	N	Y	/	Hourly energy consumption	Sequential
[61]	N	N	N	N	N	Y	/	One step energy prediction	Sequential
[62]	N	N	N	N	N	Y	/	Hourly energy consumption	Sequential
[28]	Y	Y	N	N	N	N	On/off state of system plants	1 min, 60 min and infinite time ahead energy prediction	Sequential
[48]	Y	Y	N	N	N	Y	/	1 step ahead energy prediction	Sequential
[63]	Y	Y	N	N	N	N	/	building energy consumption and peak power demand of next day	Sequential

blocks of buildings, Dawood developed a new software named Local Energy Manager to predict energy demand a day ahead. The Local Energy Manger used the algorithm of Exponentially Weighted Extended Recursive Least Square (EWE-RLS) algorithm which was developed based upon a standard Kalman filter [9]. The load of single residential is always more volatile than the aggregated load at building levels. In order to solve this problem, Estebsari et al. developed a hybrid model based on time series encoding and convolutional neural network (CNN). Their proposed method reached a mean absolute percentage of error of around 12% [10]. Coincidently, Aprillia et al. proposed an open-ended prediction method which incorporated whale optimization algorithm, discrete wavelet transform, and multiple linear regression model to handle unsteady variation of end-user load as well we steady system load [11]. Deep learning algorithms, especially long short memory (LSTM) and its adapted version were used for time series prediction and obtained satisfactory results. Khan et al. proposed a system which employed one dimensional CNN and LSTM for load forecasting and scheduling operational times of appliances. The scheme was proved to be able to save 2.223kWh of energy per day after scheduling for one day and 78.79kWh of energy per day after scheduling for one month [12]. Khalid et al. used multiple variables as input for LSTM to predict electricity demand and price, and proved that the proposed model structure has higher accuracy than conventional univariate LSTM [13]. Fan et al. integrated LSTM with human behavior patterns, and further improved the model with a multi-layer neural network [14]. However, one of defects of deep neural network is that it requires larger size of dataset than traditional algorithms which hinders its wide application. As for supply side energy prediction, Ju et al. established a short-term prediction model for photovoltaic generation. The model is developed based on self-attention mechanism and multi-task learning algorithm. The self-attention mechanism is a kind of Encoder-Decoder network for feature extraction. Their experimental contrast showed that the performance of proposed method were increased by 14.82% and 8.09% compared with CNN and LSTM [15].

2.2. Existing feature engineering research on data-driven models for HVAC energy forecasting

The data-driven models used for HVAC load forecasting can generally be categorized into two classes: univariate forecasting models and multivariate forecasting models, as shown in Fig. 1. Univariate forecasting models rely only on historical values of a time series to predict future values. They do not involve feature engineering. The representative algorithms include auto regressive, moving average, and auto regressive integrated moving average models. Several studies have shown that univariate forecasting models are less accurate than other data-driven models [16,17] because they cannot capture the relationship between the target variable and exogenous variables. Multivariate forecasting models, as the name suggests, predict future values by establishing the mapping relationship between the prediction target and multiple variables. Machine learning models belong to this category. Conventional machine learning models (such as artificial neural network, support vector machine, and random forest models) require structured data for training. Appropriate and well-engineered features can considerably improve model performance [18]. Deep learning models, as part of a broader family of artificial neural networks, can extract features from raw data and conduct dimensionality reduction automatically [19]. Feature engineering mainly includes the following two steps:

1. Finding raw features that may influence the prediction target. This step requires extensive industry experience. Apart from using directly measurable features, researchers also use new features created from raw ones. Previous studies have confirmed that using assembled features can lead to better performance than using raw features [20].
2. Reducing feature dimensionality by selecting the most prominent features or computing principal components from existing features. This step is called feature extraction. Because a large feature dimension may cause the "curse of dimensionality" [21] and because redundant features may degrade model performance [22] dimensionality reduction is a necessary procedure.

2.2.1. Raw feature identification

A machine learning model maps the relationship between the input parameters and the prediction target. It is vital to capture appropriate input parameters that are the driving factors for variations in the prediction target. Energy consumption in an HVAC system is determined by various factors, which can be classified into the following four categories:

1. outdoor weather conditions,
2. building and system characteristics,
3. the indoor environment, and
4. occupant number and activity.

Outdoor meteorological parameters, including dry-bulb temperature, relative humidity, dew point temperature, solar radiation, and wind speed, are the main factors influencing building HVAC energy consumption. Almost all studies use weather parameters as input features for predicting building HVAC energy consumption. Some studies use directly observable weather parameters, whereas others use processed parameters such as cooling degree day (CDD) and heating degree day. Using degree days is a simple method of measuring cooling or heating demand. It integrates the information of balance temperature, which indicates when the energy system should be switched on [23], thereby enabling model performance to be improved to some extent.

From the perspective of training data, there are two types of methods for developing building energy prediction models. The first involves the use of the historical energy data of the target building itself as training data, while the second involves the use of the historical data of other buildings. These two types of modeling tasks are referred to as *sequential prediction* and *parallel prediction*, respectively, in this study. The main difference between these two cases is the choice of features. For the sequential prediction task, factors in categories 2 and 3 mentioned earlier can be excluded because these factors remain unchanged over time. Variables with no or small variations are not effective as model training features [24]. Moreover, after retrofitting, a building cannot be considered the same as before. For new buildings or buildings without historical data, only the parallel prediction method can be used to estimate energy consumption. This type of prediction task is more complicated. It requires historical energy data as well as the design and operation information of various training buildings. Energy data can be easily obtained from a building management system, but information regarding building characteristics (such as system operation strategy) is difficult to collect and quantify. As can be seen in Table 1, most previous studies have focused on sequential prediction.

Building energy consumption is influenced by occupants through two factors: number of occupants and energy use behavior. The human body is a natural heat source. In addition, the energy of office equipment is highly correlated with the number of occupants. A study conducted by Wei et al. [25] suggested that the influence of occupants on the energy use in an office building is more significant than that of weather. However, the number of occupants and their activities are difficult to record. Therefore, for most cases, the time index (i.e., hour of the day, day of the week, weekday or weekends, holiday, etc.) is used instead to reflect the effect of occupants on building energy consumption.

In addition to the aforementioned features, historical energy data are often used as training features. The reason for this is twofold. The first is that a building exhibits similar energy use characteristics for identical days. The energy consumption curve of office buildings fluctuates on a weekly basis. In this regard, Fan et al. [26] used 7- and 14-day-ahead energy consumption as inputs for next-day energy demand prediction. The second reason is the thermal mass of the building. The weather conditions and energy demand of previous time steps which may influence the current energy uses are often used for short-term and small-granularity (i.e., hourly) energy prediction. Fan et al. [27] used the building cooling load, outdoor temperature, and relative humidity during the past 24 h as input features. A summary of the features used for building energy prediction is provided in Table 1.

According to our literature review, new feature creation for prediction model development has seldom been studied. Mena et al. [28] transformed numbers 1–24, which represent the hours of the day, by applying a cosine or sine function. This is a useful technique to process periodic parameters. Sha et al. used CDD instead of directly measured temperature as a model input feature [23]. It should be noted that data leakage is an important issue when selecting input features. Data leakage refers to the use of information that would not be available at the time of prediction, causing the predictive score overestimate the prediction accuracy [29]. Specifically, using future meteorological parameters as input features is a type of data leakage, but its impact can be neglected because of accurate weather forecasting. However, Ding et al. [30] used chilled water volume of next day, which cannot be obtained in advance, as input features to predict the cooling load. As a result, the predictive score may have been overestimated.

2.2.2. Feature selection

Generally, two types of feature extraction methods are used. Type I refers to the selection of certain important features from the initial feature space. Type II involves transforming the initial feature space into a new one and then selecting a subset of features from the new space. However, the physical meaning of the selected new features is difficult to interpret. The principles and characteristics of each method are listed below.

(1) Filter method

The filter method ranks the importance of each feature using information theoretic or correlation criteria and then selects a subset of high-score features [31]. The Spearman [32] and Pearson [33] coefficients are two commonly used criteria for estimating the correlation between each input feature and prediction target. However, only a linear relationship can be recognized based on these two criteria.

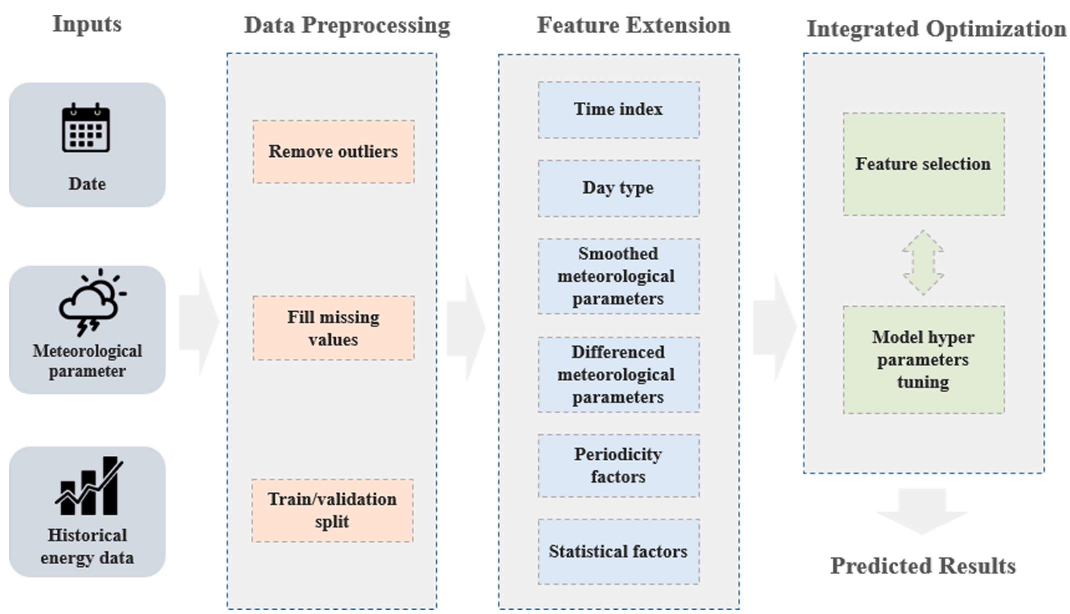

Fig. 2. Framework of multi-granularity HVAC energy forecasting toolkit.

(2) Wrapper method

The wrapper method uses predictive scores obtained via machine learning as evaluation metrics to measure all possible feature subsets and find the optimal subset [34]. The goal of the wrapper method is to find optimal feature subsets from the initial feature space. The greedy-stepwise-based [35] wrapper method employs a greedy search on the feature space and stops when adding or deleting any remaining feature does not increase the evaluation score. However, this method is highly inefficient when the feature dimension is large. Some efficient search strategies have been developed to reduce the computational complexity. Evolutionary algorithms, which are based on the concept of evolution, are commonly used as alternatives. These algorithms have been proven to be effective in finding optimal or near-optimal solutions of complex functions [36]. Aurora et al. employed two multi-objective evolutionary search algorithms, ENORA and NSGA-II, to perform feature selection [37]. Salcedo-Sanz et al. used a modified harmony search optimization algorithm to select a feature subset [38]. The Boruta algorithm is another efficient wrapper method. It performs a top-down search for relevant features by comparing the importance of original attributes with the importance achievable at random, as estimated using permuted copies of the attributes [39]. Huang et al. used the Boruta algorithm to calculate the importance of each feature and a random forest to select an appropriate subset [40]. Candanedo et al. used the Boruta package to find all relevant features from among several initial variables [41].

(3) Embedded method

The embedded method differs from the wrapper method in that it integrates a feature selection process with a model training process. Regularization and tree-based approaches belong to this category. Regularization adds a penalty to each model parameter to reduce model freedom and avoid overfitting. Jain et al. employed L1 regularization (Lasso) to the loss function and found that their method outperformed a support vector regression model without regularization [42]. Guo et al. also performed feature selection using Lasso [33]. Tree-based models (such as random forest, LightGBM, XGBoost, and CatBoost) can not only achieve outstanding predictive performance but also use feature importance for feature selection. The importance of a feature is computed as the (normalized) total reduction of the criterion resulting from that feature [43]. Yuan et al. [44] applied a random forest to select the top 10 features for heating energy prediction.

(4) Principal component analysis

In contrast to the aforementioned feature extraction methods, principal component analysis (PCA) reduces feature dimensionality by mapping original features into a lower-dimensional space whose variables are linearly uncorrelated. Ding et al. [45] employed PCA integrated with wavelet decomposition and reconstruction and with correlation analysis to obtain reasonable model inputs. Li et al. [46] analyzed the effect of PCA on building load prediction. The result showed that kernel PCA results in better performance than conventional PCA without feature selection does. Yuldiz et al. [47] discussed how to determine the dimensionality of a reduced feature space.

(5) Autoencoder

The purpose of an autoencoder is similar to that of PCA. An autoencoder is a supervised algorithm that can compress the dimensionality of input data. It is often used in deep learning networks to reduce the dimensions of large-scale data. However, autoencoders are rarely used in the field of building energy prediction because the feature space is usually small [34]. Fan et al. [48] extracted an equal number of features using different methods (including fully connected autoencoders, one-dimensional convolutional autoencoders, and generative adversarial networks) and compared the suitability of those methods for building energy prediction with that of conventional methods.

3. Framework of multi-granularity HVAC energy forecasting toolkit

As shown in Fig. 2, the HVAC energy forecasting toolkit performs three functions: data preprocessing, feature extension, and integrated optimization of feature selection and model hyperparameters. First, the raw data are preprocessed to produce clean data by removing outliers and filling in missing data in accordance with the three-sigma standard, which states that data exceeding three times of the standard deviation are outliers. In the preprocessing stage, 80% of the clean data are used as the training dataset, while the remaining are used as the validation dataset. Then, six types of high-dimensional features are extended from the raw ones. The feature extension procedure is designed to extract

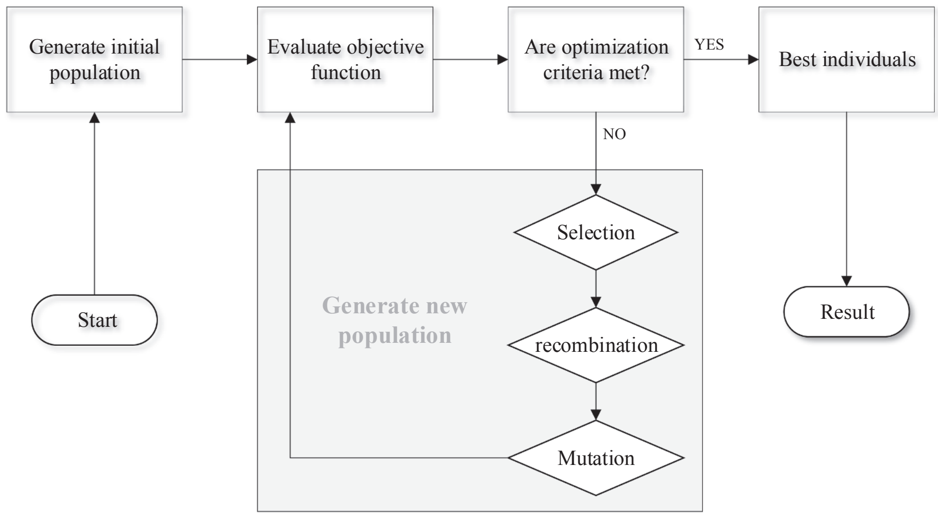

Fig. 3. Structure of a standard genetic algorithm.

hidden information within raw features so that it can be recognized by the model more easily. Given that different machine learning models are sensitive to different features, the developed toolkit adopts an integrated optimization algorithm that is based on the elitist genetic algorithm (EGA), to select appropriate features and model hyperparameters.

3.1. Feature extension

3.1.1. New features from date

Building energy consumption is highly correlated with occupant number and activity. However, traffic data are difficult to obtain for most cases. Therefore, features representing the occupant number are usually denoted by categorical features such as the hour of the day (denoted by 1–24), day of the week (denoted by 1–7), and day of the month (denoted by 1–31). In addition, occupant activity characteristics are different for different types of days (e.g., weekdays, weekends, and holidays) for most building types. Information regarding variations in energy consumption for different types of day can also be represented by categorical features denotes as 0–1. Apart from national holidays, weekends, and weekdays, there is another type of day, which is often ignored: the last working day before a holiday. Many people opt to take this day off to extend their vacation. In the feature creation framework, features of the time index (denoted by 1–7) and special type of day (denoted by 0–1) are created from the raw feature of the date to represent the impact of occupancy variation on building energy consumption.

3.1.2. New features from meteorological parameters

Outdoor meteorological parameters, especially temperature, are the driving factors for variations in building energy consumption. Therefore, directly observable temperatures, such as dry-bulb temperature and dew point temperature, are indispensable input features for energy prediction models. However, changes in the indoor environment are always delayed and are more gradual than changes in the outdoor environment because of the thermal inertia of the building. Therefore, some studies have employed meteorological parameters of previous time steps as input features. The partial autocorrelation coefficient should be calculated to determine the lagged time steps. However, for cases with large time lags (6 h for example), too many meteorological parameters of previous time steps ($4 \times 6 = 24$ if four meteorological parameters are used as input features) will be included in the model, which may cause the curse of dimensionality. In this study, we introduce smoothed meteorological parameters as model input features to mitigate the collision between lagging and feature dimensions. The smoothing method involves the use of Savitzky–Golay smoothing filters [49]. These filters are commonly used to smooth digital data. They retain data tendency and variation by removing high-frequency fluctuations. The smoothing process is known as convolution by fitting successive subsets of adjacent data points with a low-degree polynomial using the method of linear least squares [50]. In addition, the first and second differentiations of smoothed temperature data are calculated as input features to represent the temperature variation.

3.1.3. New features from historical energy data

Historical energy data have been primarily used as prediction targets instead of input features for data-driven models. This results in wastage of valuable information regarding the intrinsic variation law inside data series. In this section, historical energy data are explored to obtain more information through periodical analysis and statistical analysis.

(1) Periodical factor

Human activities tend to follow a weekly pattern. Consequently, building energy consumption exhibits a similar periodicity. In addition, the energy consumption characteristics of a building are similar for the same type of day. For example, the energy consumption of the previous Monday can be used to predict that of the next Monday. In this study, we use a unique periodical factor to characterize the energy consumption on each weekday and hour.

The periodical factor for each weekday is calculated using the following equation:

$$dr_{i,j} = \frac{\overline{e}_{i,j}}{\overline{e}_j} \tag{1}$$

where $dr_{i,j}$ is the periodical factor for each weekday in a month, with $i = 1, \cdots, 7$ (for Monday, …, Sunday, respectively), $j = 1,\ldots,12$ (for January, …, December, respectively); $\overline{e}_{i,j}$ is the mean energy consumption of the ith weekday of the jth month; and $\overline{e}_j$ and is the average value of daily energy consumption of the jth month.

The periodical factor for each hour is calculated using similar method:

$$hr_{i,j} = \frac{\overline{e}_{i,j,k}}{\overline{e}_{j,k}} \tag{2}$$

where $hr_{i,j}$ is the periodical factor for each hour of a particular weekday in a month, with $i = 1,\ldots, 24$, $j = 1,\ldots,7$, $k = 1,\ldots, 12$; $\overline{e}_{i,j,k}$ is the mean energy consumption for the ith hour of the jth weekday in kth month;

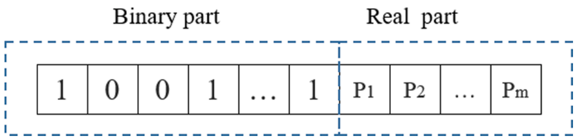

Fig. 4. Chromosome structure.

Table 2
Summary of models and hyper-parameters to be optimized.

Model name	Abbreviation	Hyperparameters
Linear regression	LR	–
Support vector regression	SVR	C, epsilon, gamma
Artificial neural network	ANN	hidden_layer_sizes, learning_rate_init
Random forest	RF	n_estimators, max_depth, min_samples_leaf, min_samples_split, max_leaf_nodes
CatBoost	CAT	–
LightGBM	LGBM	max_depth, num_leaves, n_estimators min_data_in_leaf, learning_rate

Catboost is able to great quality without parameter tuning [64].

and $\bar{e}_{j,k}$ and is the average value of hourly energy consumption of the jth weekday in kth month.

(2) Statistical factor

The statistical factors uses a time-series data vector to extract basic information using the mean, median, maximum, minimum, skew, and standard deviation of historical energy data. Features thus extracted contain statistical information of a data series. For each data point in the training dataset, the value of each temporal statistic is calculated as follows:

$$t_{i,j} = T_i\left(Y_j\right) \tag{3}$$

where T_i represents the statistical formula including the mean, median, maximum, minimum, skew, and standard deviation; j denotes the weekday (from Monday to Sunday) or the hour (from 1 to 24); and Y_j is the data series of the jth weekday. It should be noted that the order of energy consumption time series data cannot be disrupted. When calculating the periodical factors and statistical factors of the Nth data point, only the N-1 data points before the target one can be used to avoid data leakage.

3.2. EGA-based integrated optimization of feature selection and hyperparameters

A genetic algorithm is an effective optimization method that is inspired by natural selection. As illustrated in Fig. 3, a set of individuals called chromosomes is randomly initialized at the start of the algorithm. The chromosomes can be mutated and altered using crossover and mutation operators to generate a new population. The fitness of each chromosome is measured using the objective function. Compared with the standard genetic algorithm, EGA retains the few best chromosomes in the new population, which significantly improves the algorithm's performance. One percent of the best chromosomes are chosen as elite in this study. The convergence curve of an EGA is always nonincreasing. In this study, a chromosome shown in Fig. 4 is composed of two parts: a binary part and a real part. The binary part is a string of 0/1 genes with the same length as that of the total features; 1 indicates that the feature of the corresponding location is selected, and 0 indicates that it is not. The real part is a string of real genes representing model hyperparameters. The developed toolkit employs six popular machine learning models having simple and complex structures. The models and corresponding hyperparameters to be tuned are listed in Table 2. To reduce computation time, the hyperparameters that affect the model performance most significantly are selected. In order to control calculation time, the maximum iteration number of EGA is set to be 800, and the population size is 100. The probability of mutation and crossover are set to be 0.1 and 0.5 respectively.

The objective function of the EGA is the validation accuracy measured using the coefficient of variation of the root mean square error (CV-RMSE), which can be calculated using the following equation:

Fig. 5. Interface of TTBEMS.

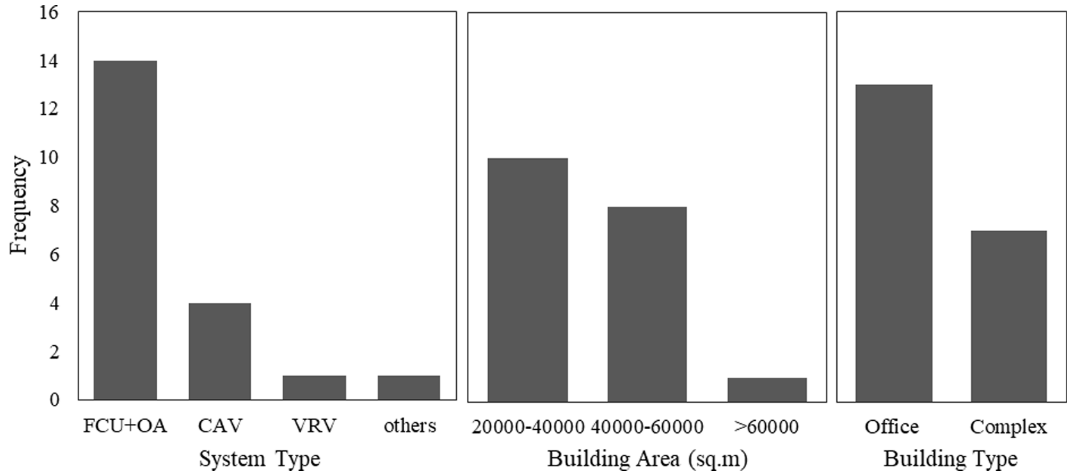

Fig. 6. Basic information of test building portfolio (FCU + OA is short for fan coil with dedicated outdoor air system. CAV is short for constant air volume system. VRV is short for variable refrigerant volume system.)

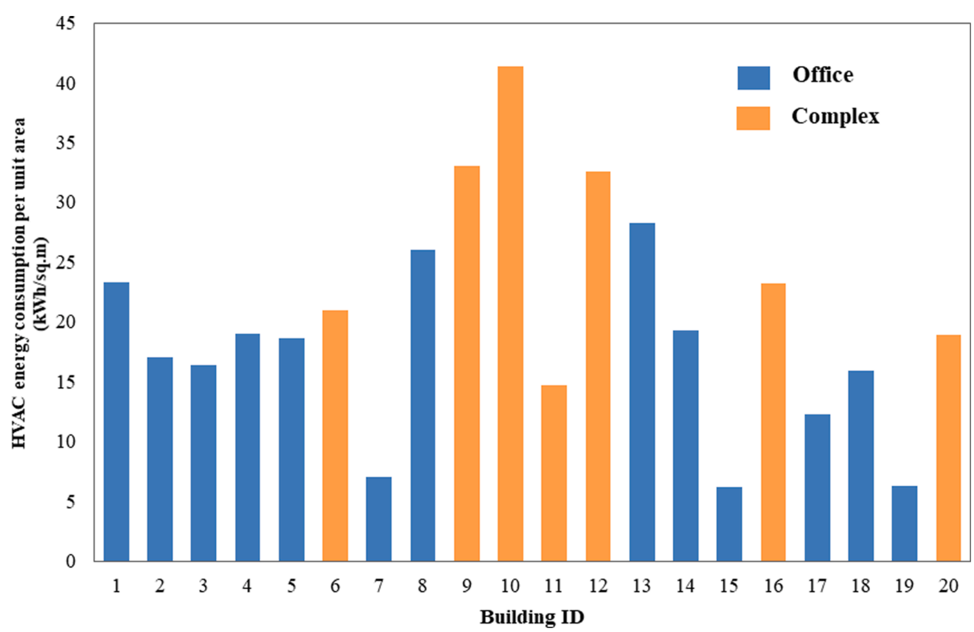

Fig. 7. HVAC energy consumption per unit area of each test building.

$$\mathrm{CV-RMSE} = \sqrt{\frac{\sum_{k=1}^{n}(y_k - \widehat{y_k})^2}{n}} \Bigg/ \frac{\sum_{k=1}^{n} y_k}{n} \tag{5}$$

where y_k is the test data value, $\widehat{y_k}$ is the predicted data value, and n is the number of test data sizes.

4. Comparative study

4.1. Data description

The performance of a data-driven model should be evaluated in terms of two parameters: accuracy and stability. A model with high stability performs well on various datasets. However, most existing studies only evaluate model accuracy based on one or a few datasets and are unable to guarantee model performance on other datasets. In this study, we use cooling energy data of summer, collected from 20 large-scale commercial buildings in central urban area of Shanghai, China for the period of May 1, 2017, to October 31, 2019, to validate the toolkit comprehensively. The energy data comes from a building energy sub-metering platform named TTBEMS. TTBEMS was sponsored by Shanghai government to collect building energy consumption data and implement energy efficiency monitoring for large commercial buildings in Shanghai central urban area. Its interface is shown in Fig. 5. This platform provides data including building meta data, energy equipment list, outdoor conditions, degree days and sub-item power consumptions. In this study, building meta data, weather conditions and power consumptions of HVAC equipment are used. The building meta data provides building basic information including building type, HVAC system type, building area and number of layers. As illustrated in Fig. 6, the test building portfolio contains 13 office buildings and 7 commercial complex buildings. The HVAC system of most test buildings comprises a fan coil with a dedicated outdoor air system. Fig. 7 illustrates the HVAC energy consumption per unit area of each building in cooling season. Evident variation in energy usage intensity can be observed for each building. The commercial complex buildings have higher energy consumption than that of office buildings. As is shown in Fig. 8, the operating patterns of office and commercial complex buildings are obviously

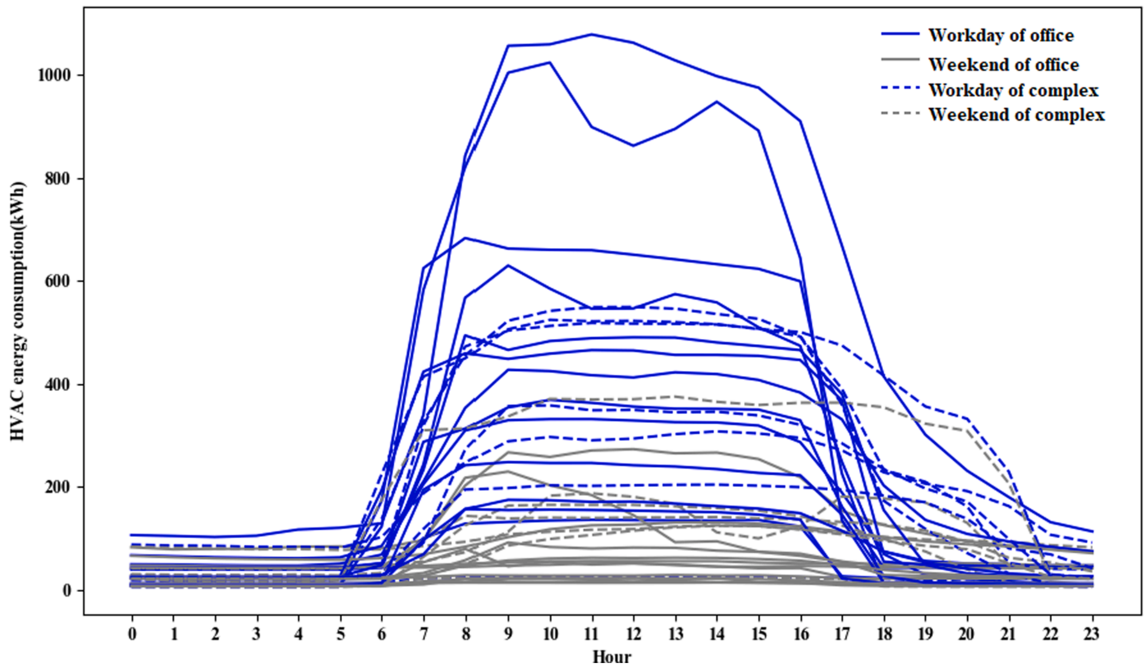

Fig. 8. Average energy patterns of each test building on workday and weekend.

Table 3
Summary of comparative baseline models.

Baseline model	Features	Model hyper-parameters
A	Extended with selection	Default
B	Extended without selection	Optimized
C	Conventional	Optimized
D (LSTM)	Conventional	Default

different. The daily energy profiles of office buildings typically rise at around 5 a.m. and fall at 6 p.m., while the latter timestamp is around 10 p.m. for commercial complex buildings. So the commercial complex buildings have much longer operating time for both workdays and weekends which resulting in higher energy intensity. The energy data for each building were recorded at an hourly rate. Therefore, each case contains 13,248 data points. Apart from directly using hourly energy data to train the models, we also aggregate the data to obtain the daily frequency (which contains only 552 data points for each case) for evaluating the model performance when the available training data size is small. Meteorological parameters recorded hourly, including dry-bulb temperature, dew point temperature, relative humidity, and wind velocity, were obtained from a local weather station installed at Hongqiao Airport.

4.2. Comparative baseline model

Table 3 lists all the comparative baseline models examined in this study. Four baseline models are developed to fully validate the toolkit proposed here. The first three baseline models are all developed based on algorithms listed in Table 2 but have different input features and hyperparameters. Model A uses selected features but default model hyperparameters. The model hyperparameters are optimized using the EGA. Model B employs all the extended features as is listed in Table 4 without selection or optimized hyperparameters, whereas model C employs the most conventional features, including raw meteorological parameters and the time index features included in Table 4. Baseline model D employs long-short term memory network which has been proved to be a superior algorithm for time series prediction. The architecture of model D is different from the first three models. LSTM, as a

Table 4
Summary of fully extended features.

Category	Feature name	Abbreviation
Raw meteorological features	Dry bulb temperature	DryT
	Dew point temperature	DewT
	Relative humidity	RH
	Wind velocity	Vel
Smoothed meteorological features	Smoothed temperature	filter_T
	Smoothed relative humidity	filter_RH
Differenced meteorological features	1st-order differenced temperature	dif_T
	1st-order differenced relative humidity	dif_RH
Time index features	*i*th day of the week	weekday
	*i*th day of the month	day
	*i*th month of the year	month
	*i*th week of the year	week
	*i*th hour of the day	hour
Day type features	If this day is weekend	is_weekend
	If this day is holiday	is_holiday
	If this day is first day of holiday	is_first_of_holiday
	If this day is last day of holiday	is_last_of_holiday
	If this day is last day of workday	is_last_of_workday
Periodicity factor features	Periodicity factor of each weekday/hour	periodicity_rate_daily/hourly
Statistics factors features	Mean of each weekday/hour	mean_daily/hourly
	Median of each weekday/hour	median_daily/hourly
	Min of each weekday/hour	min_daily/hourly
	Max of each weekday/hour	max_daily/hourly
	Skew of each weekday/hour	skew_daily/hourly
	Standard deviation of each weekday/hour	std_daily/hourly

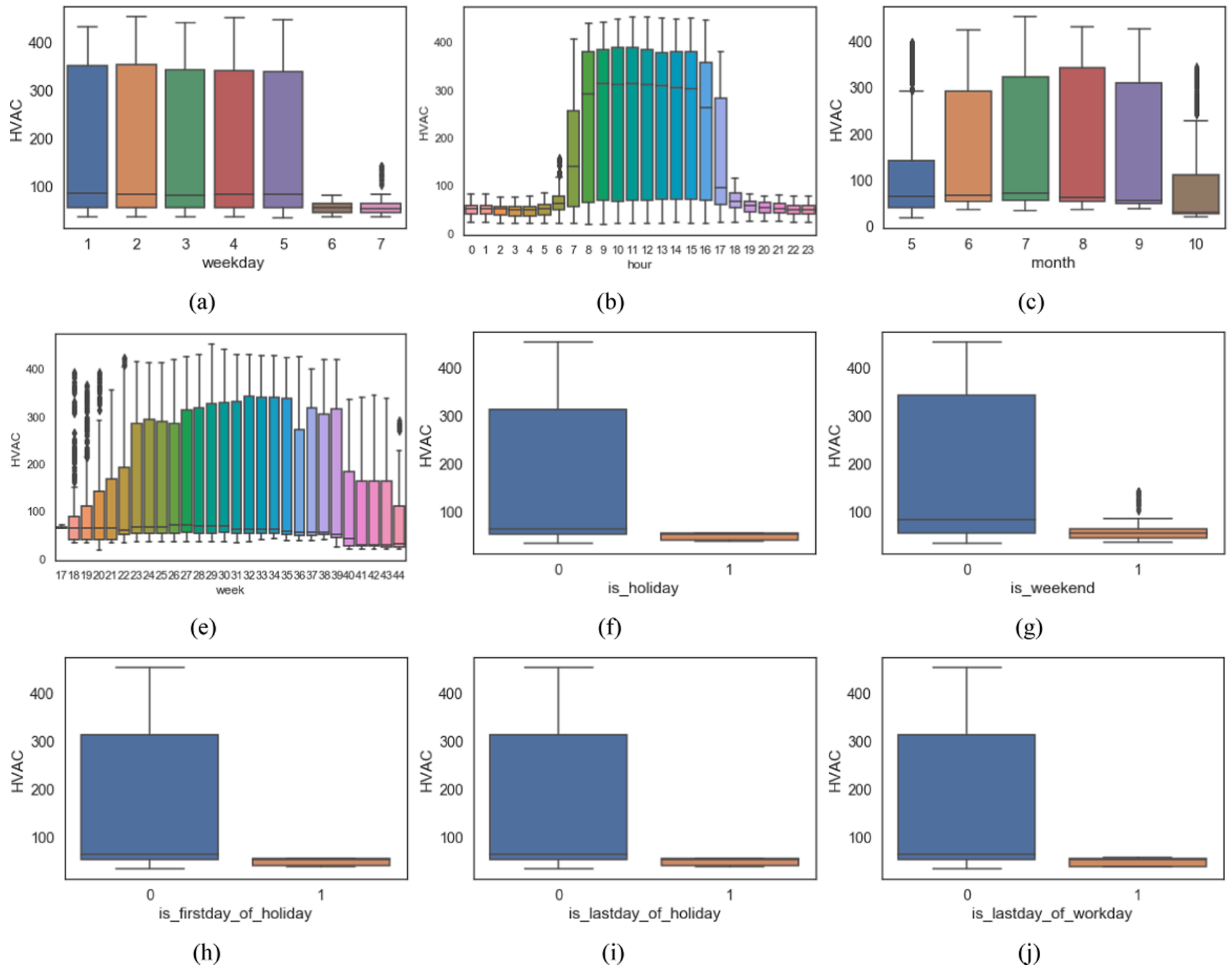

Fig. 9. Distribution of energy data extracted on each categorical feature.

sequence to sequence model, generates a sequence of output from a sequence input vectors whereas others are one to one mapping models. In this study, the LSTM is trained to use the energy data and features in the previous time step and 24-h ahead features to predict 24-h ahead energy consumption. EGA is used to optimize the following parameters: (1) number of hidden layers; (2) dropout ratio of each layer; (3) batch size. The activation function and optimizer are set to be *ReLU* and Adam respectively. The toolkit and first three baseline models are all tested using the hourly and daily datasets mentioned in the previous section. Model D is tested only using hourly datasets because the size of daily dataset is too small to train a LSTM model.

5. Results and discussion

5.1. Extended features

The fully extended features obtained using the aforementioned feature extension method are summarized in Table 4. A total of 32 and 25 features were created for developing hourly and daily energy prediction models, respectively. The time index features and day type features are categorical features, which should be specially encoded for most machine learning algorithms except for tree-based models. The other numerical features can be directly fed into the models. Boxplots were prepared as visual supplements for energy data distribution (see Fig. 9). The box length is an indicator of the data range, while the black line inside the box is the median of the dataset. If the data distribution extracted based on one category differs significantly from that extracted based on another, this categorical feature is important for model development. As dry-bulb temperature and relative humidity are the most influential meteorological parameters affecting HVAC energy consumption [23], smoothing and differencing operations are conducted only on these two parameters to reduce the computational burden of the optimization calculation. In addition, the day-type features can be further extended to include features apart from the five listed in Table 4. For example, a feature indicating whether there are conferences or social events can be created for hotel buildings whose lodging ratio would be extremely high during a conference or social event.

5.2. Evaluation of hourly and daily energy predictions

As mentioned in the previous section, model performance should be evaluated in terms of both accuracy and stability. The distribution and statistical indicators (i.e. mean, median and standard deviation) of hourly energy prediction results calculated from the target model and comparative baseline models are displayed in Fig. 10 and Fig. 11, respectively. Lower CV-RMSE mean values indicate higher prediction accuracy. The more compact the distribution, the better the stability. From a mathematical view, the prediction stability can also be represented by standard deviation values of the prediction results. The lower the standard deviation, the better the stability. It is evident from Fig. 10 that the models that use extended features have considerably higher accuracy than those that use conventional features, regardless of which machine learning algorithm is employed. Fig. 11 shows violin plots that display the hourly prediction distribution, where the width of each violin reflects the data frequency. As shown in Fig. 11, LSTM, as a kind of

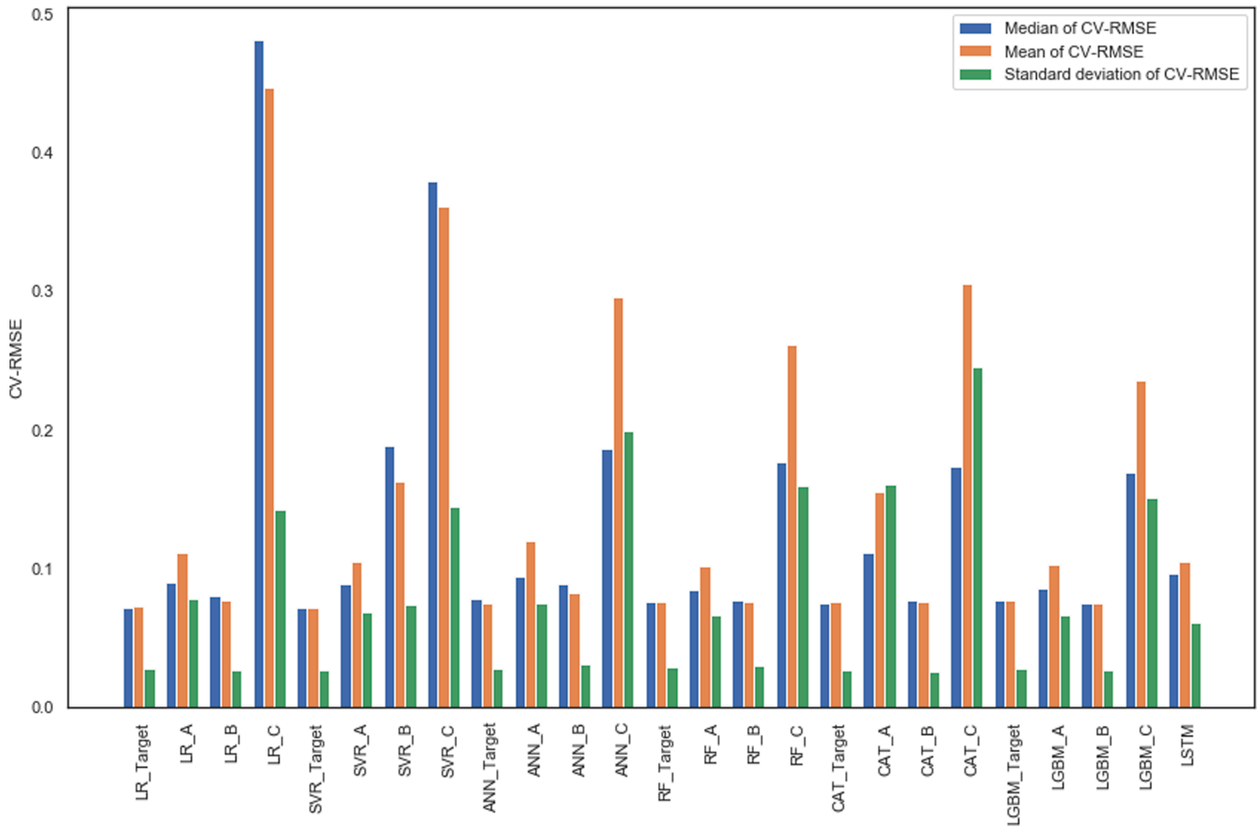

Fig. 10. Median, mean and standard deviation of hourly CV-RMSE of target model and comparative baseline models.

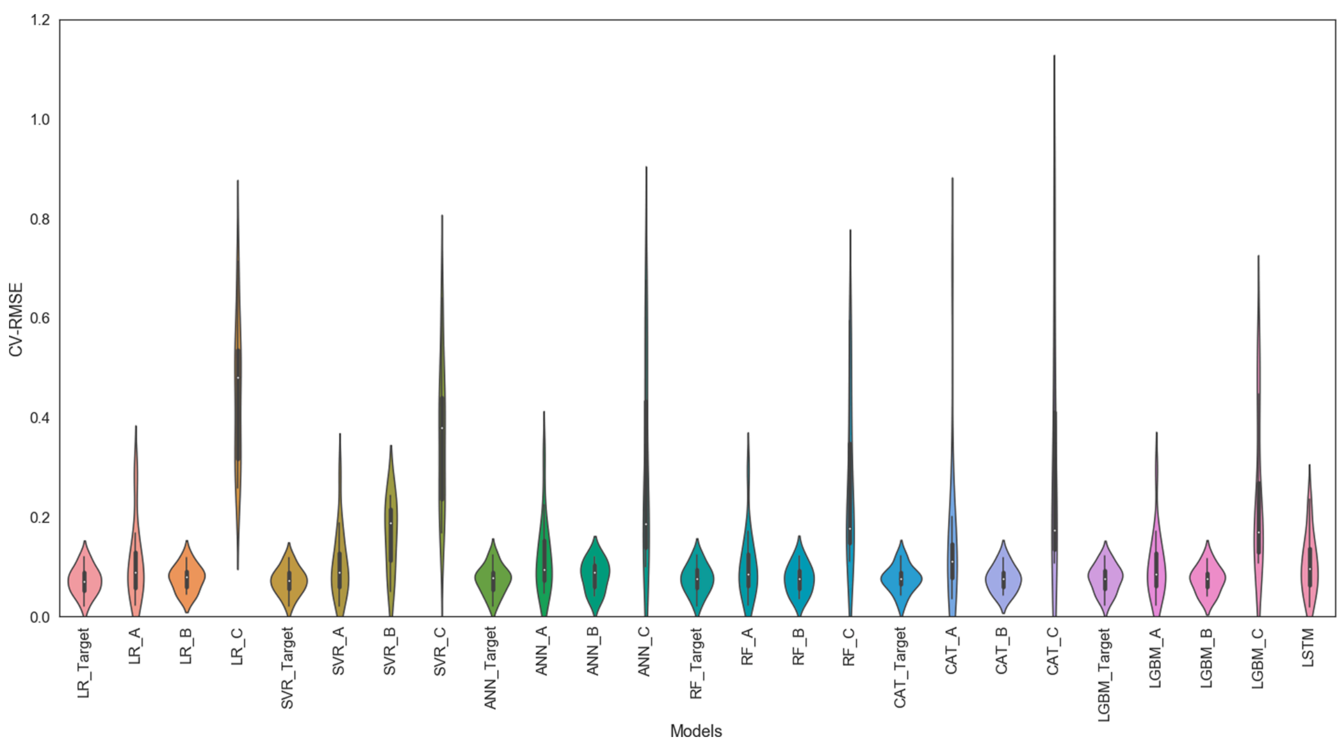

Fig. 11. Hourly energy prediction results of target model and comparative baseline models.

deep learning model, outperforms shallow learning models in terms of both accuracy and stability when using conventional features. However, the extended features are not employed to LSTM as its required training data size will be multiplied with feature size increasing. The available amounts of data in this study are not able to build high quality LSTM models using extended features. For shallow learning models, complex algorithms perform better than simple algorithms when conventional features are used to train the model; however, the deviation of such

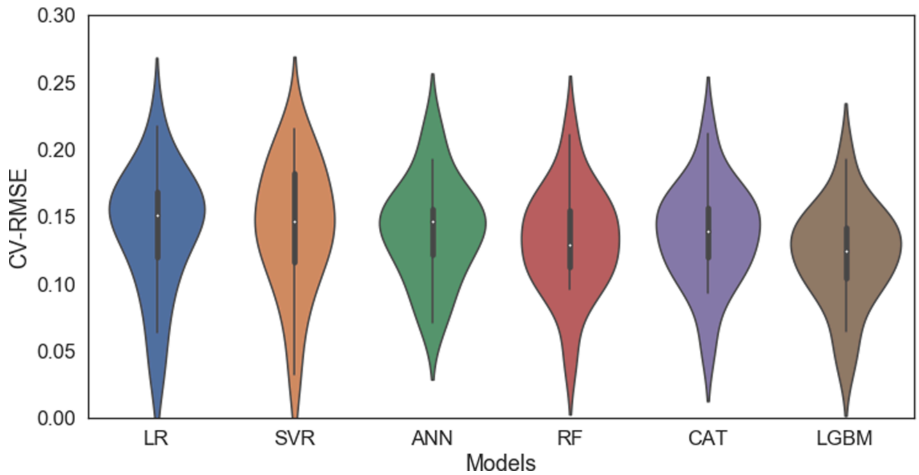

Fig. 12. Daily energy prediction results of target models.

cases is large, which indicates low stability. On the other hand, the distribution of prediction results obtained using models developed based on extended features is much more compact than that of models developed based on conventional features; this verifies that extended features improve model stability. This phenomenon indicates that well-engineered features can help reduce the dependence of a model on complex algorithms. Of all the models using extended features, the target model that employs the integrated optimization of feature selection and hyperparameter tuning exhibits the best stability among all seven regression algorithms, although it cannot always achieve the lowest CV-RMSE. Fig. 12 shows the prediction results of the target models trained using daily energy datasets. The accuracy is slightly lower than that of the models trained using hourly energy datasets but is nevertheless acceptable for engineering purposes. The model performance is degraded mainly because of the small dataset, which shrinks considerably after the hourly frequencies are aggregated to daily frequencies.

6. Conclusion

The energy forecasting toolkit proposed by this paper provides a convenient way to predict energy demand in advance which is the key to the success of demand response (DR) programs. It relies more on novel features rather than complicated algorithms to get high accuracy Feature engineering, an essential procedure for developing data-driven energy forecasting models, has rarely been investigated. In this study, we first presented a comprehensive overview of existing energy prediction and feature engineering research. The variables that are commonly employed as input features of heating, ventilation, and air-conditioning (HVAC) energy prediction models were analyzed and summarized. The advantages and disadvantages of different feature selection methods were discussed. Then, an HVAC energy forecasting toolkit that focuses on feature engineering was developed. Six types of features were extended from three types of raw features. Compared with features extracted using previous data-driven models, those extracted in this study are much more informative. Comparative experiments involving three baseline models developed using field-test data collected from 20 commercial buildings were conducted to evaluate the usefulness of the proposed toolkit. The results validate the toolkit from the perspectives of both accuracy and stability. Compared with energy prediction models developed using conventional feature sets, models developed using extended features have higher accuracy and greater stability. The target model that employs integrated optimization of feature selection and hyperparameter tuning outperformed all the comparative models. The mean coefficient of variation of the root mean square error (CV-RMSE) of the target models for hourly energy prediction was found to be $<8\%$, which means that the toolkit is acceptable for engineering applications.

CRediT authorship contribution statement

Huajing Sha: Conceptualization, Data curation, Methodology, Writing - original draft. **Peng Xu:** Supervision. **Meishun Lin:** Writing - review & editing. **Chen Peng:** Writing - review & editing. **Qiang Dou:** Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This work was supported by Terminal building energy conservation technology system and engineering demonstration (2018YFC0705004).

References

[1] Pattern Recognition and Machine Learning. J Electron Imaging 2007. doi:10.1117/1.2819119.
[2] Ali U, Shamsi MH, Bohacek M, Hoare C, Purcell K, Mangina E, et al. A data-driven approach to optimize urban scale energy retrofit decisions for residential buildings. Appl Energy 2020. https://doi.org/10.1016/j.apenergy.2020.114861.
[3] Xiao F, Fan C. Data mining in building automation system for improving building operational performance. Energy Build 2014. https://doi.org/10.1016/j.enbuild.2014.02.005.
[4] Fan C, Xiao F, Madsen H, Wang D. Temporal knowledge discovery in big BAS data for building energy management. Energy Build 2015. https://doi.org/10.1016/j.enbuild.2015.09.060.
[5] Spandagos C, Ng TL. Fuzzy model of residential energy decision-making considering behavioral economic concepts. Appl Energy 2018. https://doi.org/10.1016/j.apenergy.2017.10.112.
[6] Esnaola-Gonzalez I, Javier Diez F. Integrating building and IoT data in demand response solutions. CEUR Workshop Proc.; 2019.
[7] Kebir N, Lamallam A, Moussa A. Daily peak-based short-term demand prediction using backpropagation combined to chi-squared distribution. Int J Emerg Electr Power Syst 2020. https://doi.org/10.1515/ijeeps-2020-0098.
[8] Alduailij MA, Petri I, Rana O, Alduailij MA, Aldawood AS. Forecasting peak energy demand for smart buildings. J Supercomput 2020. https://doi.org/10.1007/s11227-020-03540-3.
[9] Daood N. Short-term prediction of energy consumption in demand response for blocks of buildings: DR-BoB approach. Buildings 2019. https://doi.org/10.3390/buildings9100221.
[10] Estebsari A, Rajabi R. Single residential load forecasting using deep learning and image encoding techniques. Electron 2020. https://doi.org/10.3390/electronics9010068.

[11] Aprillia H, Yang HT, Huang CM. Optimal decomposition and reconstruction of discrete wavelet transformation for short-term load forecasting. Energies 2019. https://doi.org/10.3390/en12244654.

[12] Khan M, Seo J, Kim D. Towards energy efficient home automation: a deep learning approach. Sensors 2020. https://doi.org/10.3390/s20247187.

[13] Khalid R, Javaid N, Al-zahrani FA, Aurangzeb K, Qazi EUH, Ashfaq T. Electricity load and price forecasting using jaya-long short term memory (JLSTM) in smart grids. Entropy 2020. https://doi.org/10.3390/e22010010.

[14] Fan L, Li J, Zhang XP. Load prediction methods using machine learning for home energy management systems based on human behavior patterns recognition. CSEE J Power Energy Syst 2020. https://doi.org/10.17775/CSEEJPES.2018.01130.

[15] Ju Y, Li J, Sun G. Ultra-short-term photovoltaic power prediction based on self-attention mechanism and multi-task learning. IEEE Access 2020. https://doi.org/10.1109/ACCESS.2020.2978635.

[16] Zhou C, Fang Z, Xu X, Zhang X, Ding Y, Jiang X, et al. Using long short-term memory networks to predict energy consumption of air-conditioning systems. Sustain Cities Soc 2020. https://doi.org/10.1016/j.scs.2019.102000.

[17] Zhao L, Liu Z, Mbachu J. Energy management through cost forecasting for residential buildings in New Zealand. Energies 2019. https://doi.org/10.3390/en12152888.

[18] Brownlee J. Discover feature engineering, how to engineer features and how to get good at it. Sept 26 2014.

[19] Schmidhuber J. Deep Learning in neural networks: an overview. Neural Networks 2015. https://doi.org/10.1016/j.neunet.2014.09.003.

[20] Chandrashekar G, Sahin F. A survey on feature selection methods. Comput Electr Eng 2014. https://doi.org/10.1016/j.compeleceng.2013.11.024.

[21] Pardalos PM. Approximate dynamic programming: solving the curses of dimensionality. Optim Methods Softw 2009. https://doi.org/10.1080/10556780802583108.

[22] Trunk GV. A problem of dimensionality: a simple example. IEEE Trans Pattern Anal Mach Intell 1979. https://doi.org/10.1109/TPAMI.1979.4766926.

[23] Sha H, Xu P, Hu C, Li Z, Chen Y, Chen Z. A simplified HVAC energy prediction method based on degree-day. Sustain Cities Soc 2019. https://doi.org/10.1016/j.scs.2019.101698.

[24] Ertugrul OF, Tagluk ME. A fast feature selection approach based on extreme learning machine and coefficient of variation. Turkish J Electr Eng Comput Sci 2017. https://doi.org/10.3906/elk-1606-122.

[25] Wei Y, Xia L, Pan S, Wu J, Zhang X, Han M, et al. Prediction of occupancy level and energy consumption in office building using blind system identification and neural networks. Appl Energy 2019. https://doi.org/10.1016/j.apenergy.2019.02.056.

[26] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Appl. Energy 2014;127:1–10. https://doi.org/10.1016/j.apenergy.2014. 04.016.

[27] Fan C, Xiao F, Zhao Y. A short-term building cooling load prediction method using deep learning algorithms. Appl Energy 2017. https://doi.org/10.1016/j.apenergy.2017.03.064.

[28] Mena R, Rodríguez F, Castilla M, Arahal MR. A prediction model based on neural networks for the energy consumption of a bioclimatic building. Energy Build 2014. https://doi.org/10.1016/j.enbuild.2014.06.052.

[29] Kaufman S, Rosset S, Perlich C. Leakage in data mining: formulation, detection, and avoidance. In: Proc. ACM SIGKDD Int. Conf. Knowl. Discov. Data Min.; 2011. https://doi.org/10.1145/2020408.2020496.

[30] Ding Y, Zhang Q, Yuan T, Yang F. Effect of input variables on cooling load prediction accuracy of an office building. Appl Therm Eng 2018. https://doi.org/10.1016/j.applthermaleng.2017.09.007.

[31] Sondhi P. Feature construction methods: a survey. Sifaka Cs Uiuc Edu 2010.

[32] Cai M, Pipattanasomporn M, Rahman S. Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques. Appl Energy 2019. https://doi.org/10.1016/j.apenergy.2018.12.042.

[33] Guo Y, Wang J, Chen H, Li G, Liu J, Xu C, et al. Machine learning-based thermal response time ahead energy demand prediction for building heating systems. Appl Energy 2018. https://doi.org/10.1016/j.apenergy.2018.03.125.

[34] Sun Y, Haghighat F, Fung BCM. A review of the-state-of-the-art in data-driven approaches for building energy prediction. Energy Build 2020. https://doi.org/10.1016/j.enbuild.2020.110022.

[35] Russell S, Norvig P. Artificial intelligence: a modern approach. 2nd ed. Prentice-Hall; 2003.

[36] Evolutionary computation for modeling and optimization; 2006. doi:10.1007/0-387-31909-3.

[37] González-Vidal A, Jiménez F, Gómez-Skarmeta AF. A methodology for energy multivariate time series forecasting in smart buildings based on feature selection. Energy Build 2019. https://doi.org/10.1016/j.enbuild.2019.05.021.

[38] Salcedo-Sanz S, Muñoz-Bulnes J, Portilla-Figueras JA, Del Ser J. One-year-ahead energy demand estimation from macroeconomic variables using computational intelligence algorithms. Energy Convers Manag 2015. https://doi.org/10.1016/j.enconman.2015.03.109.

[39] https://cran.r-project.org/web/packages/Boruta/index.html.

[40] Wang Z, Wang Y, Srinivasan RS. A novel ensemble learning approach to support building energy use prediction. Energy Build 2018. https://doi.org/10.1016/j.enbuild.2017.10.085.

[41] Candanedo LM, Feldheim V, Deramaix D. Data driven prediction models of energy use of appliances in a low-energy house. Energy Build 2017. https://doi.org/10.1016/j.enbuild.2017.01.083.

[42] Jain RK, Damoulas T, Kontokosta CE. Towards data-driven energy consumption forecasting of multi-family residential buildings: feature selection via the lasso. Comput Civ Build Eng 2014:1675–82. https://doi.org/10.1061/9780784413616.208.

[43] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu. "LightGBM: A Highly Efficient Gradient Boosting Decision Tree". Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 3149-3157.

[44] Yuan P, Duanmu L, Wang Z. Coal consumption prediction model of space heating with feature selection for rural residences in severe cold area in China. Sustain Cities Soc 2019. https://doi.org/10.1016/j.scs.2019.101643.

[45] Ding Y, Zhang Q, Yuan T, Yang K. Model input selection for building heating load prediction: a case study for an office building in Tianjin. Energy Build 2018. https://doi.org/10.1016/j.enbuild.2017.11.002.

[46] Xuemei L, Lixing D, Jinhu L, Gang X, Jibin L. A novel hybrid approach of KPCA and SVM for building cooling load prediction. In: 2010 third international conference on knowledge discovery and data mining; 2010, p. 522–26. doi: 10.1109/WKDD.2010.137.

[47] Yildiz B, Bilbao JI, Sproul AB. A review and analysis of regression and machine learning models on commercial building electricity load forecasting. Renew Sustain Energy Rev 2017. https://doi.org/10.1016/j.rser.2017.02.023.

[48] Fan C, Sun Y, Zhao Y, Song M, Wang J. Deep learning-based feature engineering methods for improved building energy prediction. Appl Energy 2019. https://doi.org/10.1016/j.apenergy.2019.02.052.

[49] Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least-squares procedures. Anal. Chem. 1964;36(8):1627–39.

[50] https://en.wikipedia.org/wiki/Savitzky%E2%80%93Golay_filter.

[51] Pan Y, Zhang L. Data-driven estimation of building energy consumption with multi-source heterogeneous data. Appl Energy 2020. https://doi.org/10.1016/j.apenergy.2020.114965.

[52] Fan C, Sun Y, Xiao F, Ma J, Lee D, Wang J, et al. Statistical investigations of transfer learning-based methodology for short-term building energy predictions. Appl Energy 2020. https://doi.org/10.1016/j.apenergy.2020.114499.

[53] Yao Y, Lian Z, Liu S, Hou Z. Hourly cooling load prediction by a combined forecasting model based on analytic hierarchy process. Int J Therm Sci 2004. https://doi.org/10.1016/j.ijthermalsci.2004.02.009.

[54] Singh P, Dwivedi P, Kant V. A hybrid method based on neural network and improved environmental adaptation method using Controlled Gaussian Mutation with real parameter for short-term load forecasting. Energy 2019. https://doi.org/10.1016/j.energy.2019.02.141.

[55] Li X, Ding L, Lǚ J, Xu G, Li J. A novel hybrid approach of KPCA and SVM for building cooling load prediction. In: 3rd Int. Conf. Knowl. Discov. Data Mining, WKDD 2010, 2010. doi:10.1109/WKDD.2010.137.

[56] Tsanas A, Xifara A. Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy Build 2012. https://doi.org/10.1016/j.enbuild.2012.03.003.

[57] Chammas M, Makhoul A, Demerjian J. An efficient data model for energy prediction using wireless sensors. Comput Electr Eng 2019. https://doi.org/10.1016/j.compeleceng.2019.04.002.

[58] Idowu S, Saguna S, Åhlund C, Schelén O. Applied machine learning: forecasting heat load in district heating system. Energy Build 2016. https://doi.org/10.1016/j.enbuild.2016.09.068.

[59] Zeng A, Liu S, Yu Y. Comparative study of data driven methods in building electricity use prediction. Energy Build 2019. https://doi.org/10.1016/j.enbuild.2019.04.029.

[60] Wang Z, Hong T, Piette MA. Data fusion in predicting internal heat gains for office buildings through a deep learning approach. Appl Energy 2019. https://doi.org/10.1016/j.apenergy.2019.02.066.

[61] Mocanu E, Nguyen PH, Gibescu M, Kling WL. Deep learning for estimating building energy consumption. Sustain Energy, Grids Networks 2016. https://doi.org/10.1016/j.segan.2016.02.005.

[62] Bedi J, Toshniwal D. Deep learning framework to forecast electricity demand. Appl Energy 2019. https://doi.org/10.1016/j.apenergy.2019.01.113.

[63] Fan C, Xiao F, Wang S. Development of prediction models for next-day building energy consumption and peak power demand using data mining techniques. Appl Energy 2014. https://doi.org/10.1016/j.apenergy.2014.04.016.

[64] https://catboost.ai/.