Energy & Buildings 186 (2019) 230–243

Contents lists available at ScienceDirect

Energy & Buildings

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

Data driven parallel prediction of building energy consumption using generative adversarial nets

Chenlu Tian[a], Chengdong Li[a,*], Guiqing Zhang[a], Yisheng Lv[b]

[a] School of Information and Electrical Engineering, Shandong Jianzhu University, Jinan, 250101, China
[b] Institute of Automation, Chinese Academy of Sciences, Beijing, 100190, China

A R T I C L E I N F O

Article history:
Received 29 July 2018
Revised 22 January 2019
Accepted 26 January 2019
Available online 28 January 2019

2010 MSC:
00-01
99-00

Keywords:
building energy consumption
deep learning
parallel prediction
Generative Adversarial Nets

A B S T R A C T

Building energy consumption prediction is becoming increasingly vital for energy management, equipment efficiency improvement, cooperation between building energy and power grid, and so on. But it is still a hard work to obtain accurate prediction results because of the complexity of the building energy behavior and the frequent undulations in the energy demand. In the building energy consumption prediction, the existing historical data are usually used to construct the traditional machine learning models and the deep learning models. However, compared with the data sets in the research domains of image recognition, speech processing and other fields, the data sets in the time series prediction of building energy consumption do not have a large quantity. Although the gray model can reduce the reliability on sufficient data, the model is difficult to develop, and it still needs detailed building information that may be lost in existing buildings. To overcome such issues, based on the parallel learning theory, we propose the parallel prediction scheme for the building energy consumption using Generative Adversarial Nets (GAN). The parallel prediction firstly makes use of a small number of the original data series to generate the parallel data via GAN, and then forms the mixed data set which includes the original data and the artificial data, and finally utilizes the mixed data to train the prediction models. To verify the proposed parallel prediction method, two experiments which adopts different kinds of data sets from two real-world buildings are conducted. In each experiment, the availability of the parallel data and the rationality of the parallel prediction model are evaluated, and detailed comparisons are made. Experimental results show that the parallel data have similar distributions to the original data, and the prediction models trained by the mixed data perform better than those trained only using the original data. Comparison results demonstrated that the proposed method performs best compared with the existing methods such as the information diffusion technology (IDT), the heuristic Mega-trend-diffusion (HMTD) method and the bootstrap method. The proposed parallel prediction scheme can also be extended to other time series forecasting problems, such as the electricity load forecasting, and the traffic flow prediction.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Energy consumption has dramatically increased in the buildings over the past decades due to the population growth, increased demand for building functions and the global climate change [1]. Chinese building energy consumption have increased 1.7 times, and the building energy consumption percentage of total energy consumption stayed relatively stable between 17.7% and 20.3% [2]. Accurate prediction of building-energy is becoming increasingly vital for energy management, equipment efficiency improvement, cooperation between building energy and power grid, and so on. The building energy consumption prediction in buildings has been widely studied via different methods. Artificial intelligence models are one of the most popularly-used methods because they do not need the detailed building and environmental parameters [1,3]. The artificial intelligence models could be divided into two categories which are the traditional artificial intelligence models and the deep learning models.

The traditional artificial intelligence models applied in building energy prediction mainly include the artificial neural network (ANN) [4], the decision tree (DT) [5], the clustering [6,7], and the support vector machine (SVM) [8,9]. Recently, Guo et al. [10] developed the energy demand prediction models via the extreme learning machine and the multiple linear regression. Wang et al. [11] proposed a homogeneous ensemble approach using the Random Forest (RF) for hourly building energy prediction. Feynman

* Corresponding author

E-mail addresses: chenlutian2017@sdjzu.edu.cn (C. Tian), lichengdong@sdjzu.edu.cn (C. Li), qqzhang@sdjzu.edu.cn (G. Zhang), yisheng.lv@ia.ac.cn (Y. Lv).

https://doi.org/10.1016/j.enbuild.2019.01.034
0378-7788/© 2019 Elsevier B.V. All rights reserved.

et al. [12] presented an SVM-based lighting energy consumption prediction model for the office buildings. Wang et al. [13] identified the similar daily operation patterns of pumps in a water-source heat pump system by clustering analysis.

Although the traditional artificial intelligence models perform well in some short-term prediction, it is still a hard work to obtain accurate prediction results because of the complexity of the building's energy behavior and the frequent undulations in demand. As an evolution of ANNs, the deep learning approach allows higher levels of abstraction, and thus increases the prediction accuracy [14]. The deep learning has been applied in the energy consumption prediction and obtained good performance. For time series prediction of building energy consumption, Mocanu et al. [14] investigated two newly developed stochastic models which are the Conditional Restricted Boltzmann Machine (CRBM) and the Factored Conditional Restricted Boltzmann Machine (FCRBM). Shi et al. [15] predicted the energy consumption of an office building using echo state networks (ESNs). Li et al. [16] proposed a hybrid model combining the stacked autoencoders (SAEs) with the extreme learning machine (ELM) for building energy consumption prediction. In this hybrid model, the SAE is used for extracting the building energy consumption features while the ELM is utilized as the predictor. Chen et al. [17] predicted the annual household electricity consumption via the ensemble learning technique. In this prediction model, the extreme gradient boosting forest and the feedforward deep networks are used as the base models which are then combined by the ridge regression. Fu [18] utilized the deep belief network and ensemble technique to forecast the deterministic cooling load for high accuracy. Li et al. [19] presented a modified deep belief network (DBN) based hybrid model which combines the outputs from the DBN model with the energy-consuming pattern for the prediction of the building energy consumption.

Both the deep learning methods and the traditional machine learning methods rely on existing historical data. To obtain good performances, the training data set must be representative and contain sufficient variety [1]. However, it may be difficult, costly and time consuming to collect such sufficiently representative and wideranging data [20], especially in the new-built buildings and the energy-saving renovation buildings. If we do not have sufficient data or the sampling data deviate from the real data distribution, the prediction accuracy will be affected. To overcome such issues as well as to improve the prediction accuracy, some grey models which combines the physical models with the artificial intelligence models were proposed. In [21], Banihashemi et al. presented a hybrid objective function of machine learning algorithms to optimize the energy consumption of the residential buildings. In this model, the building design layout and the HVAC parameters were established, and then the decision tree and the artificial neural network were utilized. In [22] Dong developed a hybrid approach combining the forward and data-driven models to forecast both the total and air-conditioning energy consumptions. Their results illustrated that the accuracy can be improved comparing with the pure data-driven methods. Although the gray models can reduce the reliability on the sufficient data, the models are still difficult to be developed, because they need detailed building information which may be lost in some existing buildings [23].

In recent years, in order to enhance the learning ability of the machine learning methods, Li and Wang [24,25] proposed the parallel learning scheme. The parallel learning is a new framework for machine learning. It combined the concepts of the descriptive learning, the predictive learning, and the prescriptive learning into a uniform framework. In this way, the parallel system can improve the learning system by self-boosting [25]. The parallel learning provided a new perspective of data, knowledge and action. The parallel learning scheme has found successful applications to the intelligent transportation systems [26–28]. A specific feature of the parallel learning is the data processing via the software defined artificial system. One issue is how to generate high quality of the artificial data.

In the past few years, some methods have been proposed for generating artificial data. In [29], Huang et al. presented the diffusion neutral networks using the information diffusion technology (IDT) to obtain the derivative samples. In [30], Li et al. proposed a mega-trend-diffusion (MTD) technique which combines the mega diffusion with the data trend estimation to generate the artificial data for flexible manufacturing system. In [31], Li et al. further combined the MTD method with the heuristic mechanism (HMTD) to generate the artificial data and then utilized the mixed data for prediction. In [32], Chen et al. combined the constrained particle swarm optimization and the triangular membership to obtain the virtual samples. Besides, some scholars utilized the bootstrap method [33] and the Monte Carlo method [34] to produce the artificial data. Recently, in order to enhance the quality of the artificial data, the Generative Adversarial Nets (GAN) [35] has been used in the parallel learning scheme.

GAN is a new hot topic in the deep learning and the artificial intelligence domains [36], and can be used to explore the potential distribution of complex data, and can also generate high-quality generated samples as an effective supplement to real data [36]. The GAN model was firstly proposed by Goodfellow et al. [35] in 2014. Being different from the traditional adversarial models, the GAN does not estimate the probability density of each sample point in the sample space directly, but it implicitly expresses the distribution of the samples by the generator via an adversarial process [37]. The GAN comprehensively considers the effect of various factors such as the change tendency, the correlation of data, and so on. Douzas et al. [38] utilized the GAN to approximate the true data distribution and generated data for the minority imbalanced data sets. Li et al. [39] proposed the category sentence generative adversarial network (CS-GAN), and utilized the model to generate category sentences to enlarge the original data set. Esteban et al. [40] proposed recurrent GAN and recurrent conditional GAN to produce the realistic real-valued multi-dimensional medical time series. Tang et al. [41] proposed an approach for programmatic data augmentation by utilizing the Auxiliary Classifier GAN. Compared with the data sets in the research domains of the image recognition, speech processing and other fields, the data sets in the time series of building energy consumption do not have a large quantity [42]. The GAN provides one efficient tool to augment the data generation and prediction of building energy consumption.

In this paper, based on the parallel learning theory, the parallel prediction scheme for building energy consumption using GAN will be proposed. The parallel prediction scheme can be divided into two stages which are the energy consumption data generation stage and the prediction model learning stage. In the first stage, the parallel prediction scheme firstly makes use of a small number of the original data series to generate the parallel data via GAN, and then forms the mixed data set through mixing the original data and the artificial data. In the second stage, the parallel prediction scheme utilizes the mixed data to train the prediction models. To verify the effectiveness and advantages of the proposed parallel prediction method, two experiments using different real-world data sets will be conducted. In each experiment, the rationality of the parallel prediction model will be evaluated, and detailed comparisons will be made. Experimental results demonstrate that the parallel data from the GAN have similar distributions to the original data, and the prediction models trained using the mixed data perform better than those trained only using the original data. Comparison results indicated that the proposed parallel prediction method has better performance than the IDT, HMTD and the bootstrap methods.

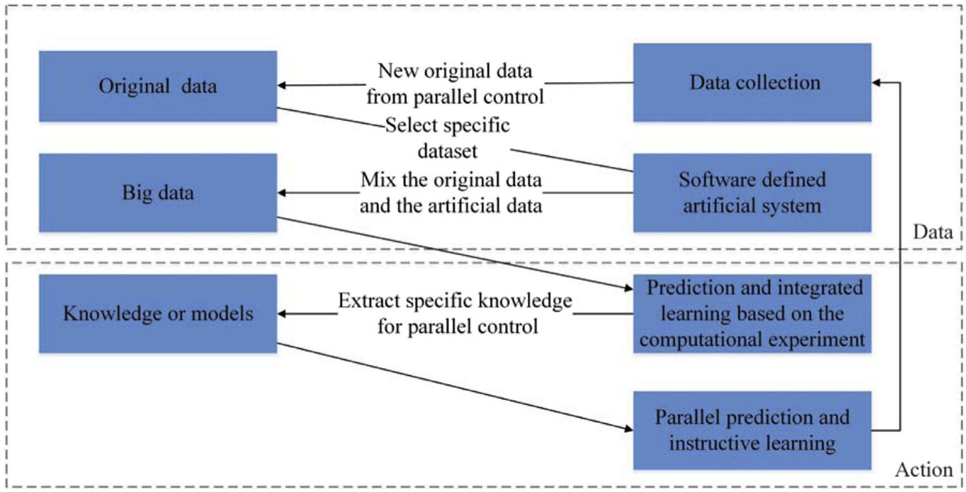

Fig. 1. The theoretical framework of the parallel learning scheme [25].

The rest of this paper is organized as follows. In Section 2, the parallel learning and GAN will be introduced, then the parallel prediction scheme will be proposed. In Section 3, the applied data and the experimental setting will be presented in detail. In Section 4, two experiments on the parallel prediction of two buildings will be performed and the comparison results will be given. Finally, conclusions of this paper will been drawn in Section 5.

2. Methodology

The proposed parallel prediction scheme for building energy consumption is based on the parallel learning architecture. For better understanding, the parallel learning will be introduced firstly in this section. Then, the GAN for data generation will be illustrated. At last, the parallel prediction scheme for building energy consumption will be proposed.

2.1. Parallel learning

The parallel learning proposed by Li and Wang [25] is a new framework for machine learning. Being different from the traditional machine learning methods, the parallel learning theory breaks the “single world outlook” and utilizes the computational experiments to conduct the predictive learning. It maps the original data to the parallel space, and then predicts and analyses the action results via a large number of simulation experiments, and finally returns the optimal action result to the reality space. The parallel learning framework is shown in Fig. 1. There are two stages in this parallel learning framework which are the data processing stage and the action learning stage.

In the data processing stage, some specific original data from the real system will be selected and input into the parallel artificial system which carries out the evolution and the iteration to generate more parallel data. In this stage, the original data are mapped to the parallel domain which reflects the similar relations as the original data.

In the action learning stage, the generated data and original data will be mixed to form a larger data set for machine learning. The machine learning model learns from the mixed data and obtains the learned knowledge. The learned knowledge means the mathematic models, data driven models, or rules, which will be used for the parallel decision or control. In this paper, the learned knowledge is the constructed data-driven models.

To some extent, the parallel learning separates the artificial system for parallel (artificial) data generation and the machine learning system for data analysis. How to generate the visual data from the artificial system is a key problem in parallel learning. Fortunately, GAN provides a good solution for this issue.

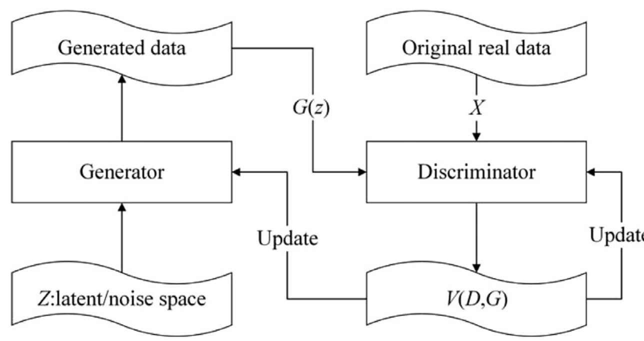

Fig. 2. Architecture of the GAN [35].

2.2. Generative adversarial nets

GAN model absorbed the idea from the game theory, and can estimate the generative models via an adversarial process [35]. The GAN is composed of two parts which are the generator and the discriminator as shown in Fig. 2. The generator is to generate new data whose distribution is similar to the original real data and make the discriminator be impossible to distinguish the generated data from it, while the discriminator is to try its best to distinguish the generated data from the generator.

In Fig. 2, X is the original data space, Z is the noise space, $G(z)$ represents a mapping from the noise space to the generated data space, and the value function $V(D, G)$ of the GAN can be described as [35]

$$V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1 - D(G(z)))], \qquad (1)$$

where x is the sample from the original space X, $p_{data}(x)$ is the distribution of the original data x, z is the noise from the noise space Z, $p_z(z)$ is defined as a prior on the input noise variables, G is a differentiable function represented by a multilayer perception, $D(x)$ describes the possibility that x comes from the real original data rather than the generator.

The generator is trained to minimize this objective, while the discriminator hopes to maximize this objective. In other words, the generator and the discriminator are trained through solving the

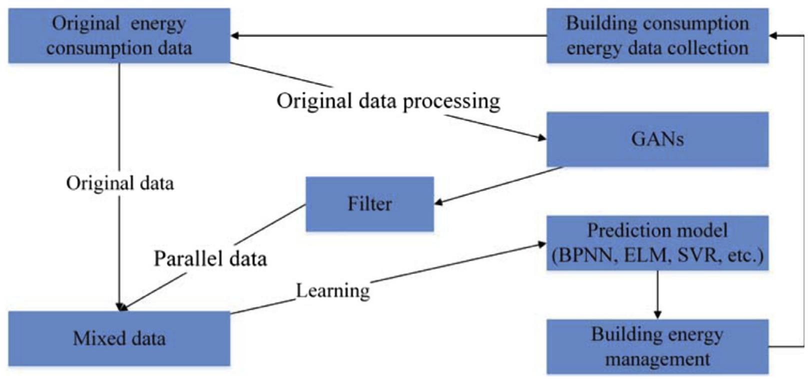

Fig. 3. Parallel prediction framework for building energy consumption.

following optimization problem [35]

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log D(x)] + E_{z\sim p_z(z)}[\log(1-D(G(z)))] \tag{2}$$

To get well-behaved GAN, the discriminator and generator play one two-step game to min-max $V(D, G)$. Firstly, the generator G is fixed and the discriminator D is optimized to maximize the discrimination accuracy. Then, the discriminator D is fixed, and the generator G is optimized to minimize the discrimination accuracy. The process continues iteratively.

In continuous space, when the generator is given, $V(D, G)$ can be described as [35]

$$V(D) = \int_x p_{data}(x)\log(D(x))dx + \int_z p_z(z)\log(1-D(G(z)))dz \tag{3}$$

$$= \int_x \left[p_{data}(x)\log(D(x)) + p_g(x)\log(1-D(x))\right]dx \tag{4}$$

where $p_g(x)$ is the generative distribution learned from the original data x. When the generator is fixed, the optimal value $D_G^*(x)$ of the discriminator can be obtained as

$$D_G^*(x) = \frac{p_{data}(x)}{p_{data}(x)+p_g(x)} \tag{5}$$

The training process of the discriminator is to maximize the log-likelihood for estimating the conditional probability of the input data from the generator or the original real data. Then the min-max game in (2) can be reformed as [35]

$$C(G) = \max_D V(D,G) = E_{x\sim p_{data(x)}}\left[\log\frac{p_{data}(x)}{p_{data}(x)+p_g(x)}\right] + E_{x\sim p_g}\left[\log\frac{p_g(x)}{p_{data}(x)+p_g(x)}\right] \tag{6}$$

The minimum value of the virtual training criterion $C(G)$ and the global optimal solution of $V(D, G)$ can be reached if only if $p_{data} = p_g$ [35].

2.3. The parallel prediction scheme for building energy consumption

The proposed parallel prediction framework for building energy consumption is shown in Fig. 3. This parallel prediction scheme is also divided into two stages which are the energy consumption data generation stage and the prediction model learning stage. In the data generation stage, the parallel prediction scheme firstly makes use of a small number of the original data series to generate the parallel data, and then forms the mixed data set through mixing the original data and the artificial data. In the second stage, the proposed method utilizes the mixed data to train the prediction models by the machine learning methods such as the back-propagation neural network (BPNN), the extreme learning machine (ELM), and support vector regression (SVR), etc.

Suppose that we have the real building energy-consumption time series data of N_1 days, which can be expressed as

$$\left\{x_t = [x_1^t, x_2^t, \cdots, x_n^t]^T\right\}_{t=1}^{N_1} \tag{7}$$

where t represents the day number, n is the number of sampling data in each day.

The parallel data generation function can be seen as a state-transition function and can be realized by the GAN. Some data generated from GANs may not be very fitted to the original data series. To solve the problem, a filter is used to remove the irregular data series. After the parallel data is filtered, the mixed data of the original and parallel data are utilized to train the prediction model.

Suppose that the GAN has generated the building energy consumption time series data of N_2 days, which are expressed as

$$\left\{y_t = [y_1^t, y_2^t, \cdots, y_n^t]^T\right\}_{t=1}^{N_2}. \tag{8}$$

Then the mixed data series can be obtained as

$$x_1^1,\cdots,x_n^1,\cdots,x_1^{N_1},\cdots,x_n^{N_1},y_1^1,\cdots,y_n^1,\cdots y_1^{N_2},\cdots,y_n^{N_2}. \tag{9}$$

For simplicity, we denote this mixed data series as

$$s_1, s_2, \cdots, s_K \tag{10}$$

in which $K = n*(N_1+N_2)$. This mixed data series will be used to construct the prediction model by machine learning methods.

In order to construct the prediction model, assume that p data in the mixed data series before time k are used to predict the value at time k. In other words, we utilize $s_{k-1},\cdots,s_{k-p}$ to predict the value of s_k by the following model

$$\hat{s}_k = f(s_{k-1},\cdots,s_{k-p}) \tag{11}$$

where $f(\cdot)$ represents the prediction model that can be realized by the machine learning methods, e.g. the BPNN, ELM, and SVR used in this study.

3. Applied datasets and experimental setting

In this section, we will firstly present the two datasets applied in the parallel prediction of building energy consumption. Then, the comparative methods and evaluation indices will be presented. Finally, the experimental procedure and settings will be introduced in details.

3.1. Applied dataset

3.1.1. Dataset for experiment 1

The first dataset is from a retail building located in Fremont, CA. The original building energy consumption dataset could be obtained from the website: https://trynthink.github.io/buildingsdatasets/. The energy consumption data was collected every 15 minutes, and 34939 samples were obtained. We aggregated the collected data into the data intervals of 30 minutes, finally we got 17469 samples.

3.1.2. Dataset for experiment 2

The second dataset applied is from the Easton Centre located in Beijing. The Easton Centre is a new-built commercial office building covering an area of 11,400.028 square meters. The building is composed of two towers: the north tower and the south tower. The consumption energy data is collected at the intervals of one hour. Totally, we have collected the houly energy consumption of this building for 240 days.

In our following experiments, we selected energy consumption data of the fourth floor in the south tower. The collected data was firstly preprocessed and the irregular data was deleted. Finally, we obtained 1944 samples.

In both experiments, to generate the parallel data of energy consumption data, the original data need to be normalized. In this study, we use the following equation to normalize the original data as

$$\bar{x}_i^t = \frac{x_i^t - \min_{i,t} x_i^t}{\max_{i,t} x_i^t - \min_{i,t} x_i^t} \tag{12}$$

where $i = 1, \cdots, n,\ t = 1, \cdots, N_1,\ \bar{x}_i^t$ means the normalized data, and x_i^t means the original data.

3.2. Comparative methods

In order to verify the advantages of the proposed method, the information diffusion technology (IDT) [29], the Heuristic-Mega-trend-diffusion (HMTD) method [30], and the bootstrap method [33] are utilized as the comparative methods in this paper. These three methods have been widely studied and applied to deal with insufficient data.

3.2.1. The IDT method [29]

The IDT method is utilized to deal with the small-sample problem in non-linear prediction. This method maps the observed real data into fuzzy sets to fill the gap and improve the prediction accuracy. The method has been widely studied and applied in many domains [29,43]. The new artificial data of the IDT method are obtained by the following equations

$$\bar{x}_i = x_i - \sqrt{-2h_x^2 \ln \phi(w)} \tag{13}$$

$$\bar{y}_k = y_k - \sqrt{-2h_y^2 \ln \phi(w)} \tag{14}$$

where $\bar{x}_i$ is the new input artificial data, and x_i is the i-th original data in the input prediction data series, $\bar{y}_k$ is the new output artificial data, and y_k is the k-th original data in the output data series. h_x and h_y are called the diffusion coefficients which can be computed as $h = \theta(b - a)$ in which a and b respectively the minimum and maximal values of x_i or y_k, θ is a variable which is suggested to be 0.6841. And, $\phi(w)$ can be determined by the values of w which measures the strength of the relationship between the paired x and y. In this paper, $\phi(w)$ is set to be 0.999999.

3.2.2. The HMTD method [31]

The HMTD method combines the Mega-trend-diffusion(MTD) method [30] with the heuristic mechanism. In the MTD method, the lowest value (L) and the maximum value (U) of the artificial data are obtained as follows

$$L = \begin{cases} CL - \frac{N_L}{N_L+N_U} * \sqrt{-2 * \hat{s}_x^2 * \ln(10^{-20})}, & L \le min \\ min, & L > min, \end{cases} \tag{15}$$

$$U = \begin{cases} CL + \frac{N_L}{N_L+N_U} * \sqrt{-2 * \hat{s}_x^2 * \ln(10^{-20})}, & U \ge max \\ max, & U < max, \end{cases} \tag{16}$$

where max and min are the maximum and minimum values of the data series, $CL = (min + max)/2$, $\hat{s}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ in which n is the number of the original data and $\bar{x}$ is the average value of the original data, N_L is the number of the data being smaller than CL, N_U is the number of the data being greater than CL.

Then, generate the temporary data t_d based on a uniform distribution between L and U. Further, compute the occurrence possibility as follows

$$MF = \begin{cases} \frac{t_d - L}{CL - L}, & t_d < CL \\ \frac{U - t_d}{U - CL}, & t_d > CL \end{cases} \tag{17}$$

Consequently, the HMTD method utilizes one heuristic algorithm to judge whether the generated data can be kept through comparing the occurrence possibility with one randomly seed within the interval [0,1]. For more details, please refer to [31].

3.2.3. The bootstrap method [33]

The bootstrap method is a statistical re-sampling method. The basic idea of bootstrapping is that inference about a population from sample data can be modelled by resampling the sample data and performing inference about a sample from the resampled data [33]. In this method, the size of the samples for sampling each time is always equivalent. The new data via bootstrap will make up the artificial data for the original data.

In this paper, to realize the validity of the bootstrap, the data at the same period of each day are integrated into a dataset for resampling. Then, the new artificial data is circularly re-sampled from the integrated dataset.

3.3. Evaluation Indices

In this study, three evaluation indices are adopted to evaluate the performances of different methods. The adopted indices are the mean absolute error (MAE), the mean absolute percentage error (MAPE), and the pearson correlation coefficient r.

The MAE is the average absolute value of the deviation between all of the individual predicted values and the real values, and can be computed as

$$MAE = \frac{1}{K}\sum_{k=1}^{K} |\hat{s}_k - s_k| \tag{18}$$

where $\hat{s}_k$ is the predicted value, s_k is the real value, K is the number of training or testing data. Smaller value of MAE indicates better performance of the prediction model. The MAE is sensitive to the large errors.

The MAPE describes the accuracy of the prediction by comparing the residual of the real value and the predicted value with the real value, and can be computed as

$$MAPE = \frac{1}{K}\sum_{k=1}^{K} \frac{|\hat{s}_k - s_k|}{s_k} \times 100\%. \tag{19}$$

MAPE expresses the accuracy in percentage. Again, the smaller the MAPE value is, the better the prediction model performs.

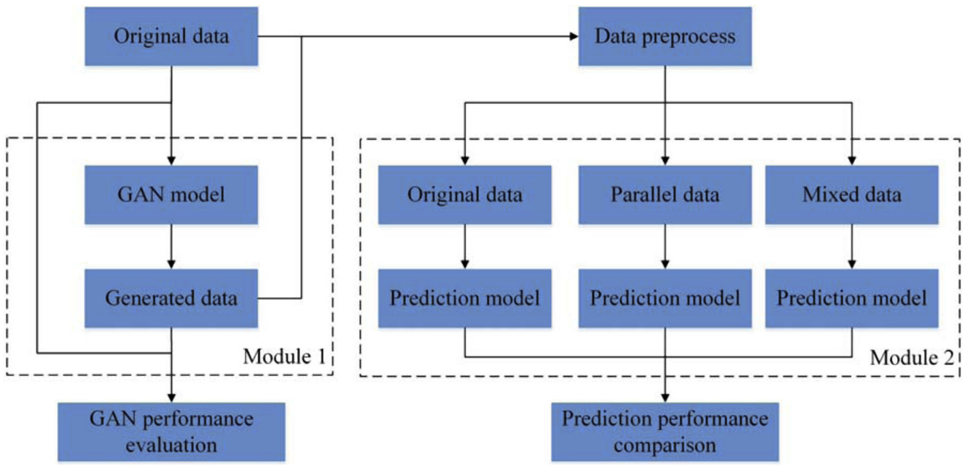

Fig. 4. The following experimental setting.

The pearson correlation coefficient is used to measure the correlation of the real value and the predicted value, and can be calculated as

$$r = \frac{\sum_{k=1}^{K}(\hat{s}_k - E(\hat{s}_k))(s_k - E(s_k))}{\sqrt{(\hat{s}_k - E(\hat{s}_k)^2)}\sqrt{(s_k - E(s_k)^2)}} \tag{20}$$

where $E(\hat{s}_k)$ means the average value of $\hat{s}_k$, and $E(s_k)$ means the average value of s_k. The value of r ranges from -1 to 1. Greater value of r expresses the better performance of the prediction model.

3.4. Experimental setting

Our experiments contain two modules as shown in Fig. 4. Module one is to verify the availability of the parallel data, while the second module aims to evaluate the rationality of the parallel prediction model.

In the first module, the generator of the GAN is a fully connected multilayer neural network. In this fully connected multilayer neural network, the activation function is tanh, and the Adam algorithm is utilized to optimize the gradient decent of the generator. The discriminator takes use of Long Short-Term Memory(LSTM), and the Adam algorithm is also used to train the discriminator. The application of these two deep learning models in the GAN improves the probability to find the hidden distributions. The processed original data series and the random noise data will be input to the GAN. Then, the GAN will be trained to generate the parallel energy consumption data series. In this module, the performance of the GAN and the availability of the parallel energy consumption data series will be evaluated in our experiments.

In the second module, the parallel data series will be mixed with the original data series. The mixed data series will adopted as the new training data sequence. All of the original, parallel and mixed data series will be used to train the prediction models separately. The same testing data set will be adopted to evaluate the prediction performances of different models.

In the second module, the basic prediction model we selected is the back-propagation neural network (BPNN). The BPNN has been widely used in the prediction field and behaves well in lots of applications [44]. Of course, the other machine learning methods can also be selected. In this paper, the ELM, SVR are also utilized as the predictors to validate the performance of the parallel prediction. The ELM [45] is a single hidden layer feed forward neural network whose parameters in the input layer and hidden layer are determined randomly while its weights between the hidden layer and the output layer are directly computed by the least square method. This learning scheme makes the ELM have fast learning speed and excellent approximation performance. Support vector regression (SVR) [46] is one special kind of support vector machine (SVM) [47]. To obtain the optimal decision function, it maps the input variables to high dimensional linear feature space by kernel functions. It has been proven that the SVR performs well in dealing with the high-dimensional features.

4. Experiments

In this section, two experiments will be conducted. And in each experiment, we will firstly verify the availability of the generated data from the GAN, and then the parallel prediction performance will be examined and compared. Finally, comparison analysis and comprehensive discussions will be given.

4.1. Retail building experiment

4.1.1. Parallel data generation for retail building

The retail building data is divided into 13 small data sets. Each small data set consists of 28 daily data sequences, and there are 48 points in each data sequence. As a result, there are 1344 data points in each small data set. We randomly selected 8 small data sets for data generation separately. In other words, we run the experiment for data generation 8 times.

In this parallel data generation process, the generator of the GAN is set to have two hidden layers with 48 nodes in each layer. And, the discriminator is set to be one hidden layer with 48 nodes while its input layer is set to have 48 nodes and the output layer is set to have one node.

The changing traces of the G-net and D-net loss in the parallel data generation for the retail building are shown in Fig. 5. In this figure, the loss represents the ability that the discriminator can distinguish the input data from the generator or the original data. For the generator, the Gloss reflects the ability that the discriminator can distinguish the data from the generator as fake, and the Dloss is defined as the loss of the discriminator which indicate that the discriminator distinguishes the data from the real data set as the fake.

Fig. 6 shows us the 24h data series of different days. From this figure, we can observe that the building energy consumption has the daily mode, as the energy consumption in each day has the similar distribution. We use the following equation to generate the daily energy consumption mode of this retail building

$$x_{eq} = \frac{1}{N_1}\left[\sum_{t=1}^{N_1} x_1^t, \sum_{t=1}^{N_1} x_2^t, \cdots, \sum_{t=1}^{N_1} x_n^t\right]^T \tag{21}$$

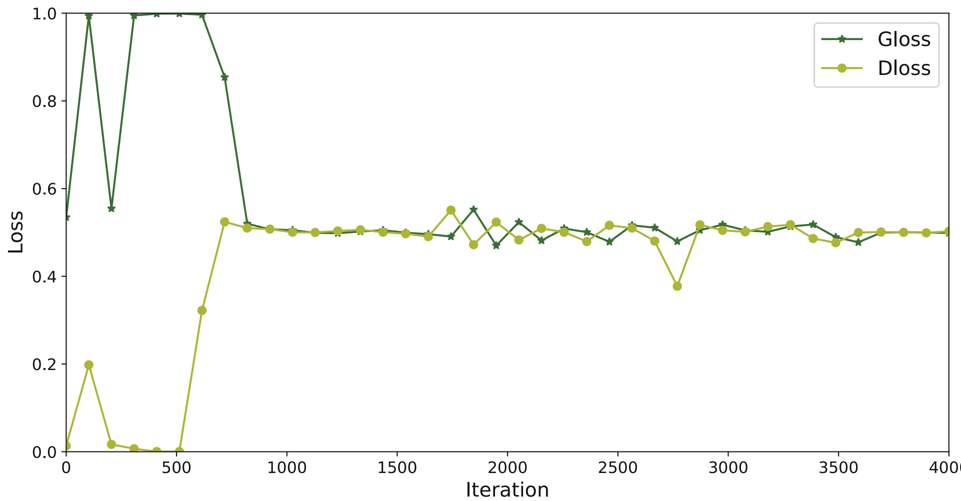

Fig. 5. The changing traces of the G-loss and D-loss in the retail building experiment.

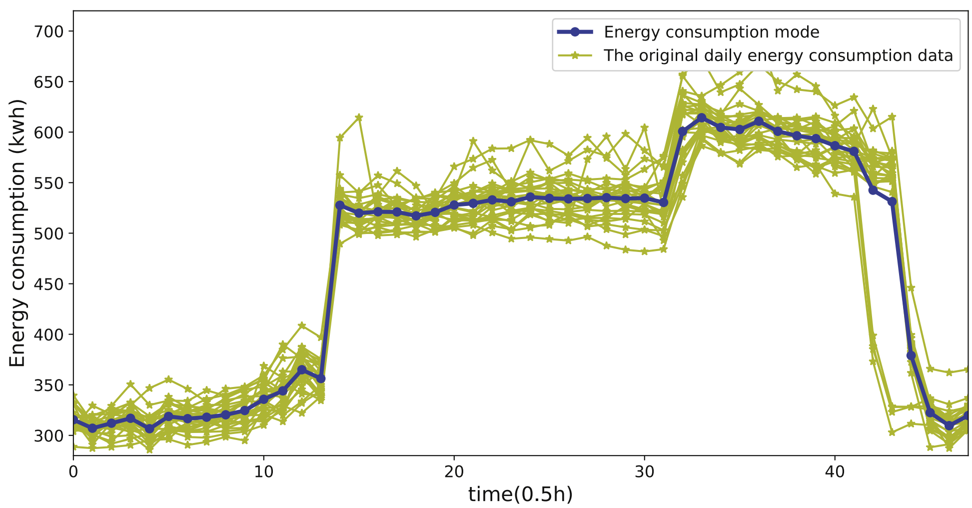

Fig. 6. The daily energy consumption mode in the retail building.

where N_1 is the number of the sampling days, and x_{eq} can reflect the daily energy consumption data distribution in the retail building.

The generated data series is compared with the daily mode x_{eq} of the original data series in the retail building. When the learning iteration of the GAN increases from 1 to 4000, the changing traces of the values of the MAE, MAPE and r for this comparison are shown in Fig. 7.

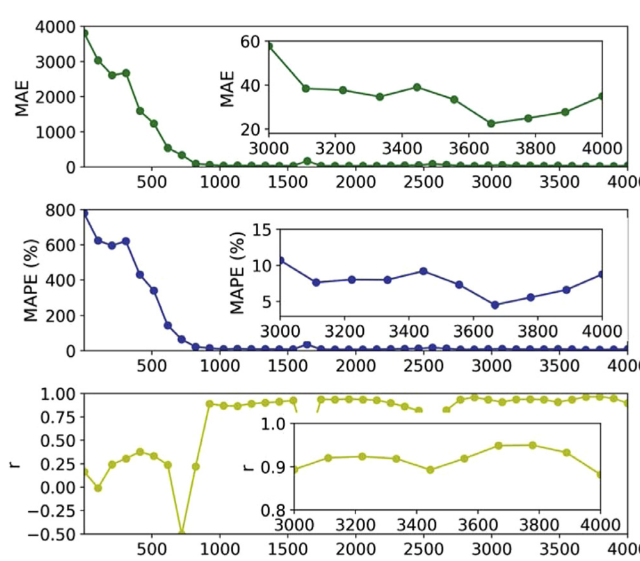

Fig. 7. The changing traces of the values of MAE, MAPE, and r for the comparison between the generated and the real retail building data when the learning iteration increases.

4.1.2. Parallel prediction of the retail building energy consumption

In this experiment, three cases will be conducted to validate the effectiveness and advantages of the parallel prediction. In the first case, the BPNN is chosen to be the predictor, and we compare the prediction performances of the original, parallel and mixed data driven prediction models constructed by the BPNN method. In the second case, to show the advantages of the proposed method for artificial data generation, the GAN will be compared with the IDT, HMTD, and bootstrap methods. In this case, the mixed data from these four methods are used to train the BPNN prediction models. Then, their performances will be compared. In the third case, the BPNN prediction model are replaced by the ELM and SVR models. In this case, the BPNN, ELM and SVR prediction models will be constructed by the mixed data from the GAN, IDT, HMTD, and bootstrap methods. Again, their performances will be compared.

Table 1
Prediction performances of the BPNN predictors constructed by the original, parallel and mixed data in the retail building experiment.

Num	Data	MAE	MAPE (%)	r	Num	Data	MAE	MAPE (%)	r
	Original	25.18	5.70	0.939		Original	30.30	6.62	0.919
1	Parallel	23.21	5.38	0.942	5	Parallel	27.77	6.04	0.919
	Mixed	24.66	5.50	0.940		Mixed	27.55	5.99	0.924
	Original	39.74	8.26	0.936		Original	26.14	5.35	0.946
2	Parallel	33.01	7.44	0.922	6	Parallel	37.02	7.29	0.931
	Mixed	28.96	6.29	0.942		Mixed	22.39	4.64	0.948
	Original	37.16	7.51	0.942		Original	41.02	7.52	0.925
3	Parallel	32.07	6.10	0.941	7	Parallel	38.56	6.83	0.915
	Mixed	34.27	6.73	0.930		Mixed	35.19	6.47	0.930
	Original	38.13	6.98	0.936		Original	32.28	5.93	0.952
4	Parallel	42.89	7.67	0.909	8	Parallel	33.39	5.97	0.937
	Mixed	38.30	6.17	0.938		Mixed	26.59	4.84	0.953
	Original	33.74	6.69	0.937					
Average	Parallel	33.50	6.59	0.927					
	Mixed	28.98	5.83	0.940					

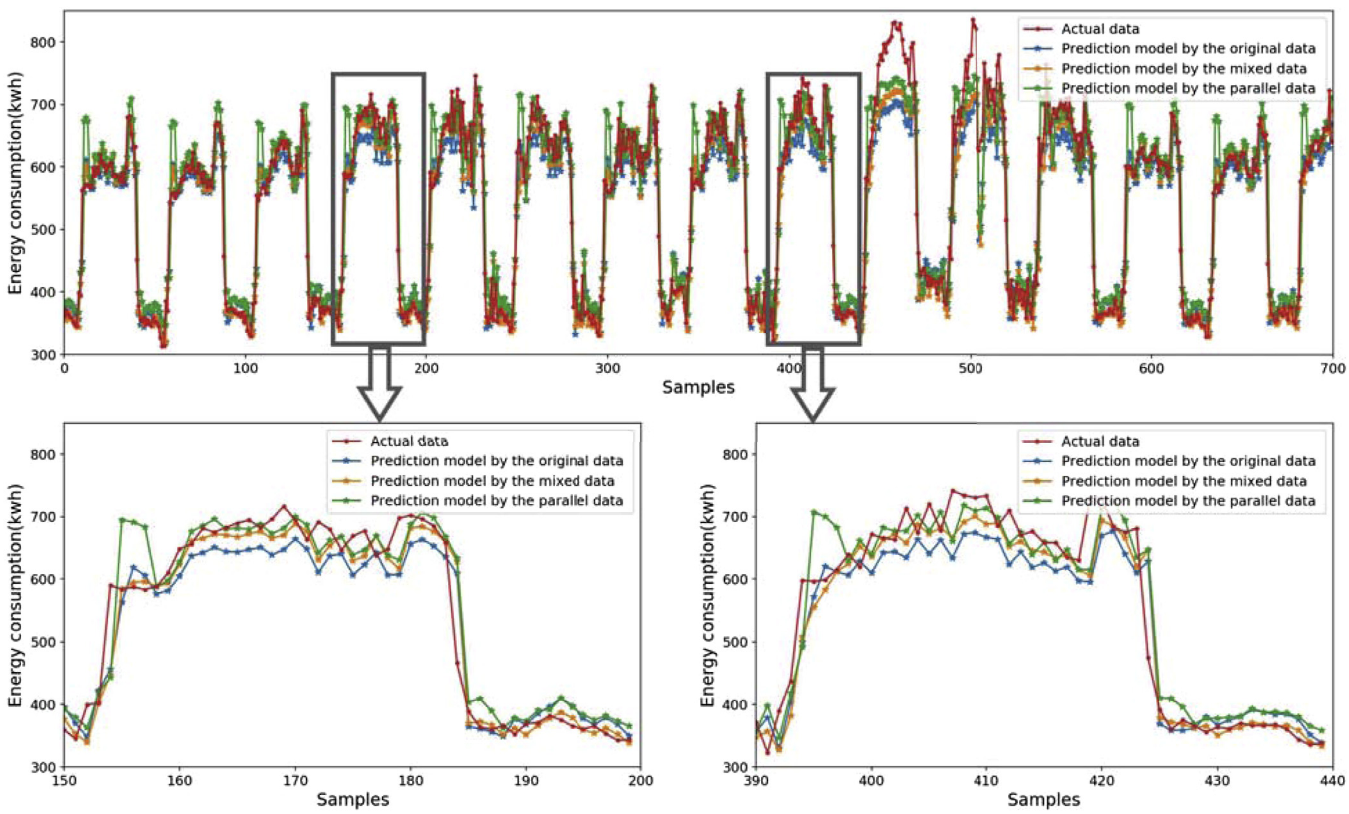

Fig. 8. Prediction results of the retail building energy consumption using the original, parallel and mixed data.

(1) Case one

In this case, three experiments are made. In the first one, there are 1344 original data points used for training. In the second one, there are 1344 parallel data for training. And, in the third experiment, the mixed data are used for training. We use the same original data set to test the models. Each experiment is run 8 times.

In each experiment, the BPNN prediction model is configured as follows: the number of the nodes in the input layer, hidden layer and output layer are respectively set to be 4, 48 and 1, and the learning rate is set to be 0.1.

The performances of the three kinds of data driven BPNN predictors are evaluated via the indices – MAE and MAPE, and r. The experimental results in the 8 runs are listed in Table 1. For better visualization, parts of the predicted results from the BPNNs trained respectively by the original, parallel and mixed data are demonstrated in Fig. 8.

(2) Case two

In order to investigate the superiority of the GAN in dealing with the parallel data generation, we conduct the second case. In this case, firstly we randomly select two data sets of the retail building as the training and testing data sets. Then, we utilize the GAN, IDT, HMTD and bootstrap methods to generate artificial data. Thirdly, we combine the same original training data with the parallel data from the four data generation methods separately to form the mixed data sets as the new training data sets. At last, the newly formed data sets are utilized to train the BPNNs, and their performances are examined via the same testing data set.

We conduct the experiment ten times for each method. The values of the MAE, MAPE and r of the BPNN models constructed

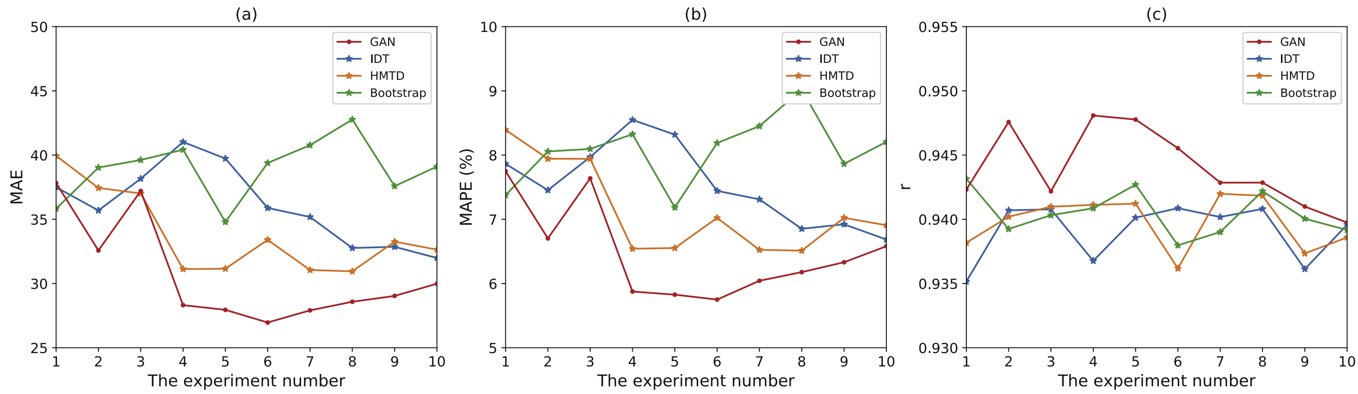

Fig. 9. Prediction performance comparison of the BPNN prediction models constructed by the mixed data from the GAN, IDT, HMTD, and bootstrap methods in the retail building experiment.

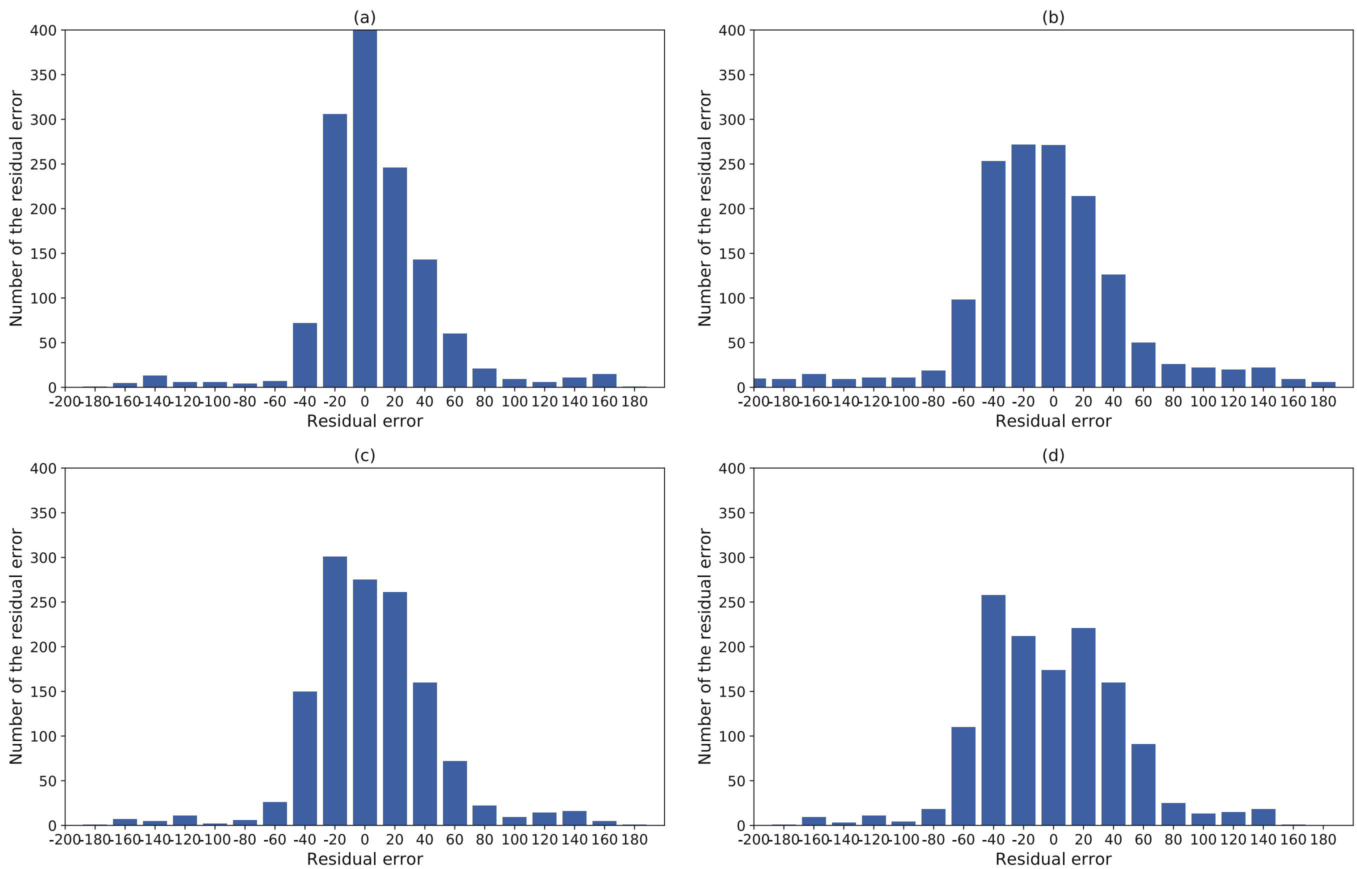

Fig. 10. Prediction error histograms of the four BPNN prediction models respectively constructed by the mixed data from the four artificial data generation methods in the retail building experiment: (a) GAN; (b) IDT; (c) HMTD; (d) bootstrap.

by the mixed data from the four data generation methods are shown in Fig. 9. Besides, for better visualization, the prediction error histograms of different data generation models are presented in Fig. 10.

(3) Case three

The third case is carried out to further verify the effectiveness and advantages of the parallel prediction model. In this experiment, the mixed data from the four data generation methods are respectively used to construct the BPNN, ELM and SVR predictors. For the ELM, the number of nodes in the hidden layer is set to be 24, while the RBF function is chosen to be the activation function. For the SVR, the penalty coefficient is set to be 0.5, and the radial basis function is chosen as the kernel function.

Again, the predictors are trained using different number of mixed training data for ten times, and in each time, the mixed data utilized for prediction is composed of the same number of the original data and various number of the parallel data. We compute the averages of the MAE, MAPE, and *r* for each prediction model, and their results are illustrated in Table 2.

4.2. Easton centre energy consumption experiment

4.2.1. Parallel data generation for easton centre

The Easton Centre data is divided into two data sets: 50% for training and the left 50% for testing. The training data set consists of 980 data points. There are 81 daily data sequences in the

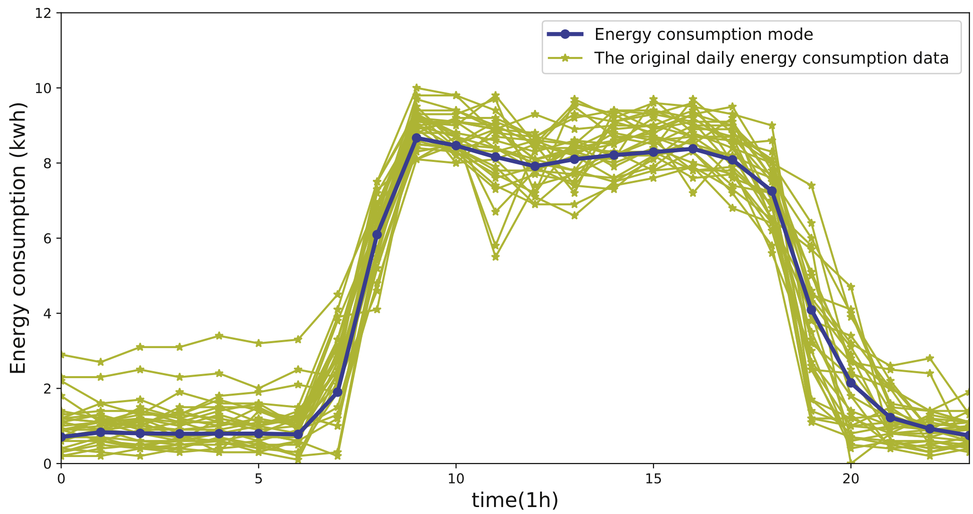

Fig. 11. The daily energy consumption mode in the Easton Centre.

Table 2
The averages of the indices MAE, MAPE and r of the BPNN, ELM, SVR using the four data generation methods in the retail building experiment.

	Method	BPNN	ELM	SVR
Average of MAE	GAN	30.63	22.10	38.68
	IDT	36.08	22.85	40.91
	HMTD	33.80	22.85	41.89
	bootstrap	38.93	23.71	38.84
	Original	40.06	24.81	40.53
Average of MAPE (%)	GAN	6.47	4.67	8.45
	IDT	7.54	4.82	8.93
	HMTD	7.14	4.85	9.18
	bootstrap	8.08	5.04	8.60
	Original	8.34	5.45	8.89
Average of r	GAN	0.944	0.948	0.942
	IDT	0.939	0.946	0.939
	HMTD	0.940	0.946	0.943
	bootstrap	0.940	0.944	0.943
	Original	0.936	0.932	0.941

training data set, and each data sequence contains 24 points. The daily energy consumption mode of the Easton Centre is demonstrated in Fig. 11. Again, we can observe that the building energy consumptions in the Easton Centre have the similar distributions in different days.

In the generation process of parallel data, the generator of GAN is set to have 24 nodes in each layer. The discriminator is set to have one hidden layer, and there are 24 nodes in the input and hidden layers and one node in the output layer.

The changing traces of the G-net and D-net loss in the parallel data generation for the Easton Centre are shown in Fig. 12. The generated data series is compared with the daily energy consumption mode x_{eq} of the original data series of the Easton Centre. When the learning iteration of the GAN increases from 1 to 4000, the changing traces of the values of the MAE, MAPE and r for this comparison are shown in Fig. 13.

4.2.2. Parallel prediction of the easton centre energy consumption

In this experiment, three similar cases as those in the retail building experiment are also considered.

(1) Case one

In this case, the parallel data of the original training dataset is generated via the GAN for ten times, and three experiments are taken into account. The first experiment utilizes 980 original data for training. The second experiment generates 980 parallel data for training. The third experiment uses the mixed data for training. The same original testing data set is used for testing the models. The prediction models are still constructed by the BPNN. The configuration of BPNN is the same as that in the retail building experiment.

We evaluate the performances of the BPNNs trained by the original, parallel and mixed data respectively. Ten experimental results are averaged and listed in Table 3. Additionally, the predicted results of these BPNNs are demonstrated in Fig. 14.

(2) Case two

In this case, 50% of the Easton Centre data is for training and the left 50% is for testing. The GAN, IDT, HMTD and bootstrap

Table 3
Prediction performances of the BPNN predictors trained by the original, parallel and mixed data respectively in the Easton Centre experiment.

Num	Data	MAE	MAPE (%)	r	Data	MAE	MAPE (%)	r
1	Parallel	0.83	24.50	0.931	Mixed	0.56	12.61	0.954
2	Parallel	0.74	20.27	0.939	Mixed	0.56	13.51	0.953
3	Parallel	0.79	21.55	0.934	Mixed	0.56	13.55	0.954
4	Parallel	0.71	18.33	0.938	Mixed	0.55	13.15	0.954
5	Parallel	0.69	18.51	0.945	Mixed	0.55	13.52	0.954
6	Parallel	0.85	25.62	0.938	Mixed	0.56	13.54	0.955
7	Parallel	0.79	22.41	0.934	Mixed	0.56	14.38	0.954
8	Parallel	0.89	27.09	0.935	Mixed	0.56	14.02	0.955
9	Parallel	0.74	17.34	0.931	Mixed	0.56	13.42	0.955
10	Parallel	0.75	20.71	0.938	Mixed	0.57	14.60	0.955
*	Parallel Average	0.79	21.63	0.936	Mixed Average	0.56	13.63	0.954
*	Original	0.67	19.42	0.952				

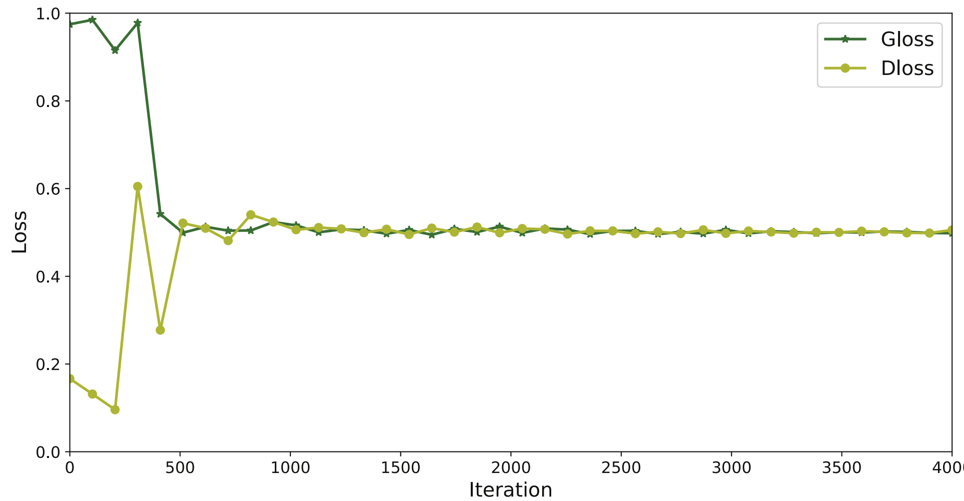

Fig. 12. The changing traces of the G-loss and D-loss in the Easton Centre experiment.

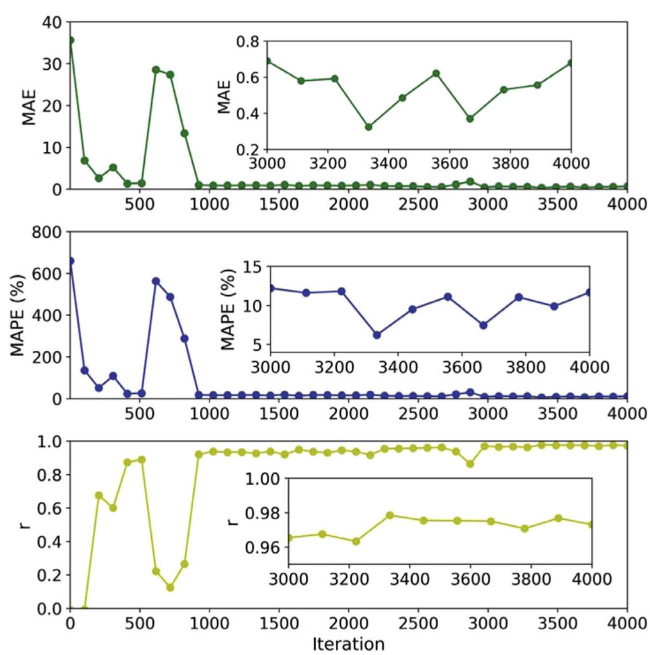

Fig. 13. The changing traces of the values of MAE, MAPE, and r for the comparison between the generated and the real Easton Centre data when the learning iteration increases.

methods are utilized to generate the artificial data. Then the artificial data from each method is combined with the same original dataset to form the newly mixed training datasets. At last, the four mixed training datasets are used to train the BPNN models.

Again, the experiment for each method is carried out for ten times. The values of MAE, MAPE and r of the four methods are presented in Fig. 15. And, the prediction error histograms of the four data generation methods are shown in Fig. 16.

(3) Case three

Again, in this experiment, the third case is carried out to further verify the effectiveness and advantages of the parallel prediction model. In this case, the mixed data from the four data generation methods are respectively used to construct the BPNN, ELM and SVR predictors. The predictors are trained using different number of mixed training data for ten times too. The averages of the values of MAE, MAPE, and r according to the three prediction models are given in Table 4.

Table 4
The averages of the indices MAE, MAPE and r of the BPNN, ELM, SVR using the four data generation methods in the Easton Centre experiment.

	Method	BPNN	ELM	SVR
Average of MAE	GAN	0.559	0.586	0.704
	IDT	0.645	0.615	0.729
	HMTD	0.609	0.604	0.731
	bootstrap	0.616	0.597	0.728
	Original	0.666	0.635	0.730
Average of MAPE (%)	GAN	13.53	14.87	20.49
	IDT	17.49	15.49	21.69
	HMTD	16.46	15.57	21.57
	bootstrap	17.15	15.12	21.47
	Original	19.10	16.76	21.78
Average of r	GAN	0.954	0.952	0.944
	IDT	0.952	0.950	0.943
	HMTD	0.953	0.951	0.943
	bootstrap	0.954	0.952	0.943
	Original	0.952	0.949	0.943

4.3. Comparison analysis and discussion

Figs. 5 and 12 illustrate the changing traces of the G-net and D-net loss in the parallel data generation for the Retail Building and the Easton Centre. We can see from these figures that with the increase of the iteration number, the losses of the G-net and D-net will gradually tend to be 0.5. This fact means that the G-net and D-net have formed a balanced confrontation situation, and the D-net can't distinguish the generated data from the G-net or the original dataset.

Figs. 7 and 13 show the comparisons between the generated data series and the daily energy consumption modes of the retail building and the Easton Centre respectively when the learning iterations increase. These figures demonstrate that the distributions of the parallel data series are similar to the modes x_{eq}. In other words, the GAN can generate effective artificial data. Taking the index r for example, in both experiments, the values of r converge to the range [0.9, 1]. This observation indicates that the GAN performs well in the distribution expression and the generation of the energy consumption data.

Tables 1 and 3 demonstrate that the values of MAE, MAPE and r of the BPNN predictors trained by the mixed data are better than those trained only by the original data. In the retail building experiment, the MAE of the mixed data driven BPNN models can be improved averagely 14.1% compared with the original data driven

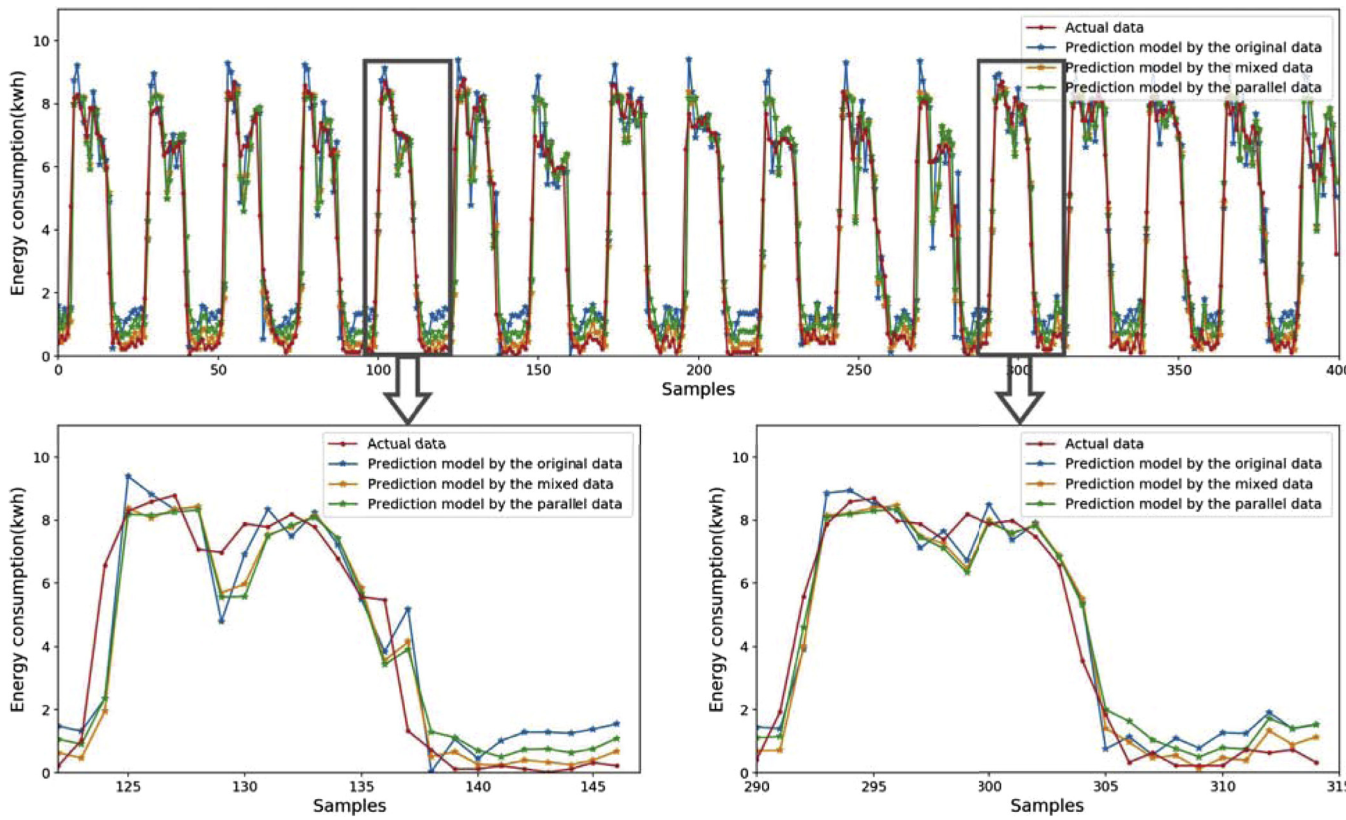

Fig. 14. Prediction results by the predictors trained using the original, parallel and mixed data in the Easton Centre experiment.

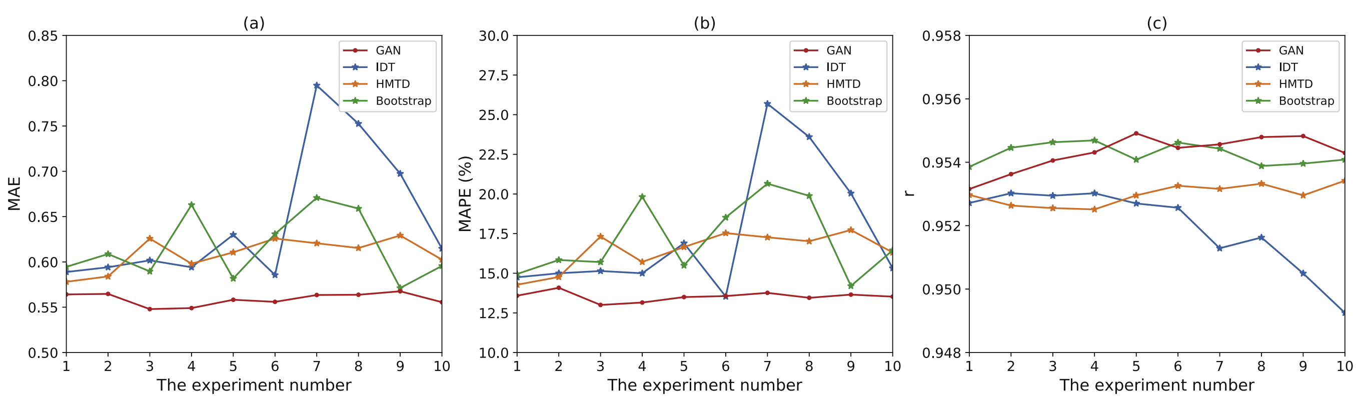

Fig. 15. Prediction performance comparison of the BPNN prediction models constructed by the mixed data from the GAN, IDT, HMTD, and bootstrap methods in the Easton Centre experiment.

BPNN models, while the improved percentage for the Easton Centre experiment is averagely 16.4%. This fact implies that the parallel data provides an excellent supplement to the original data. Although the performances of the prediction models trained only by the parallel data are not better than those using the original data, but their differences are small and acceptable. This verified the validity of the parallel data in another respect. These judgements can also be observed and verified visually from Figs. 8 and 14.

Figs. 9 and 15 present the performance comparisons of the prediction models trained by the mixed data from different artificial data generation methods including the GAN, IDT, HMTD and bootstrap methods. From these figures, we can see that there exist the unstable phenomenon in the IDT, HMTD and bootstrap methods, but the GAN has the most accurate and stable performance. Under the criteria of MAE, the GAN performs 30.96% better than the MTD, about 19.27% better than the TTD, and about 33.17% better than the bootstrap method in the first experiment, while it can be 29.11% better than the IDT, 12.44% better than the HMTD, and 17.18% better than the bootstrap method in the second experiment. The prediction error histograms in Figs. 10 and 16 can also reflect the prediction performances of the predictors trained by the mixed data from the GAN, IDT, HMTD and bootstrap methods. For the prediction error histograms, more prediction errors around 0 means better prediction performance. Again, from both figures, we can observe that the proposed method performs best. However, the GAN generates the parallel data by more learning iterations than the other comparative methods. Consequently, the computation time of the GAN is relatively higher than the IDT, HMTD and bootstrap methods.

Tables 2 and 4 are used to further display the effectiveness and advantages of the proposed method through constructing the BPNN, ELM, SVR predictors based on the mixed data generated

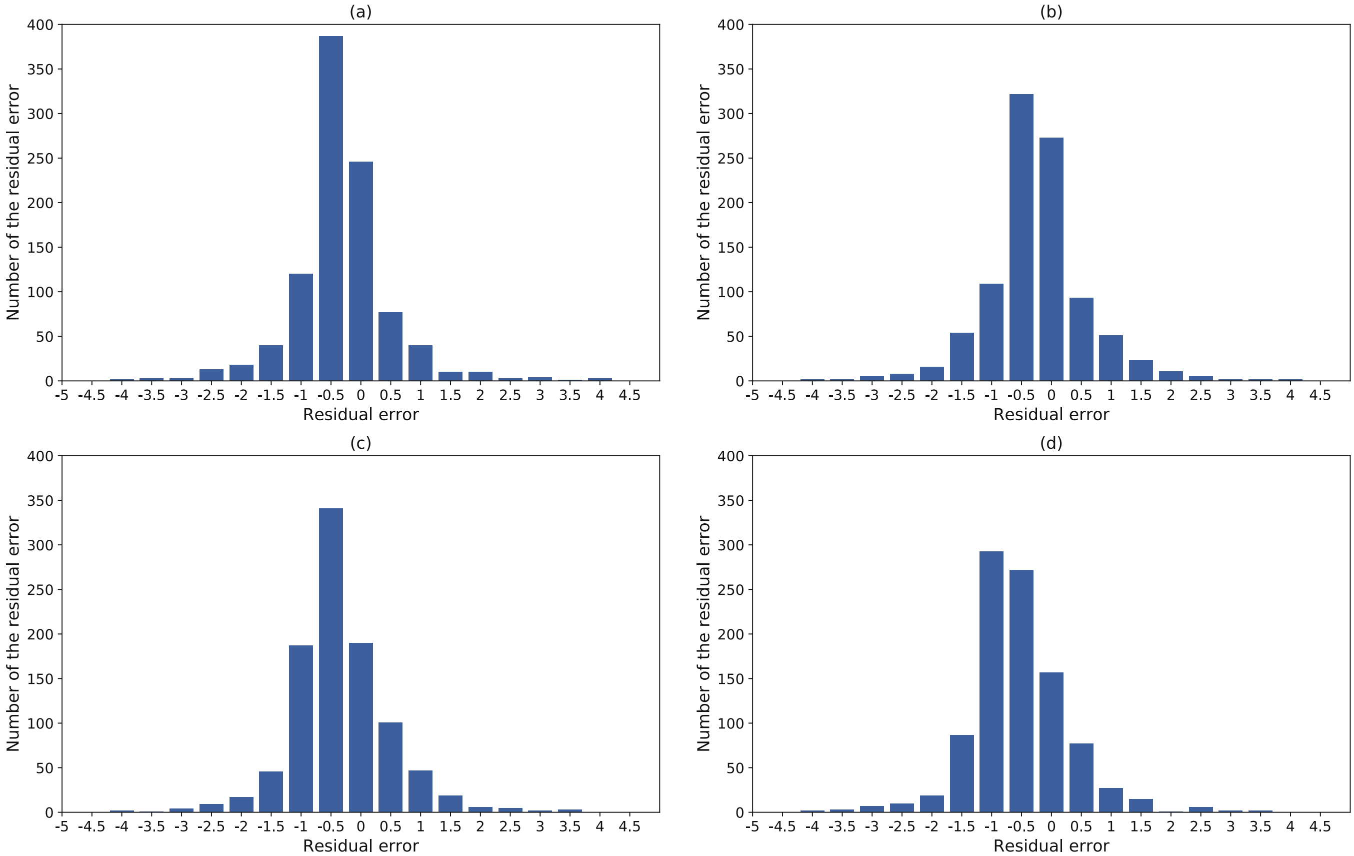

Fig. 16. Prediction error histograms of the four BPNN prediction models respectively constructed by the mixed data from the four artificial data generation methods in the Easton Centre experiment: (a) GAN; (b) IDT; (c) HMTD; (d) bootstrap.

by the GAN, IDT, HMTD and bootstrap methods. Once again, the results in both tables verify that the BPNN, ELM and SVR constructed by the mixed data from the GAN perform best than those constructed by the mixed data from the other three comparative methods.

5. Conclusion

In order to realize accurate energy consumption prediction and to solve the insufficient data problem, based on the parallel learning theory, this paper presented one parallel scheme for the building energy consumption prediction problem. The proposed parallel method for building energy consumption prediction includes two stages. The first stage utilizes the GAN to generate the artificial data (parallel data) as the supplement to the original data, while the second stage adopts the machine learning methods, e.g. BPNN, SVR, ELM etc. and the mixed data to construct the predictors for building energy consumption prediction. In other words, the GAN is used to generate the artificial data based on the real building energy consumption data, rather than being used as the predictor. In order to verify the performance and advantages of the proposed method, two detailed experiments and comprehensive comparisons were conducted. In each experiment, we considered three cases: (1) in case 1, in order to show the usefulness of the generated parallel data, the prediction performances of the BPNN based predictors trained respectively by the original, parallel and mixed data were given; (2) in case 2, in order to show the advantages of the proposed method for artificial data generation, the GAN was compared with the IDT, HMTD, and bootstrap methods. In this case, the mixed data from these four methods are used to train the BPNN prediction models, and then, their performances were compared; (3) in case 3, to further verify the effectiveness and usefulness of the generated parallel data, the BPNN, ELM and SVR based predictors trained by the mixed data from the GAN were compared with those trained by the mixed data from the IDT, HMTD, and bootstrap methods respectively.

Experiments and comparisons demonstrated that: 1) the predictor trained by the mixed data performed better than those trained only by the original data. This result implies that the parallel data generated by the GAN can reflect the running laws of the building energy consumption data; 2) compared with the IDT, HMTD, and bootstrap methods, the predictor trained by the mixed data from GAN had the best performance; 3) All the BPNN, ELM and SVR predictors which were trained by the mixed data from GAN performed better than those trained by the mixed data from the other data generation methods. Once again, this result shows the effectiveness and advantages of the proposed method. In conclusion, the proposed parallel prediction scheme can improve the prediction accuracy and can reduce the reliability on the historical real data.

In our future study, we will investigate other schemes or improved GAN architecture to make it more proper to generate the parallel building energy consumption data, and to further improve the parallel prediction accuracy. Besides, the prediction performance using larger but less incoherent data set may not be better than the prediction using coherent but not large data set. Our future work will also consider how to generate the parallel data when we meet the coherent or incoherent data set.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (61573225, 61473176), and by the Taishan Scholar Project of Shandong Province (TSQN201812092). The authors would like to thank the reviewers for their insightful comments that have helped to improve this paper.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.enbuild.2019.01.034.

References

[1] K. Amasyali, N.M. El-Gohary, A review of data-driven building energy consumption prediction studies, Renewable and Sustainable Energy Reviews 81 (2018) 1192–1205.
[2] T. Huo, H. Ren, X. Zhang, W. Cai, W. Feng, N. Zhou, X. Wang, China's energy consumption in the building sector: A statistical yearbook-energy balance sheet based splitting method, Journal of Cleaner Production 185 (2018) 665–679.
[3] C. Li, J. Yi, H. Wang, G. Zhang, J. Li, Interval data driven construction of shadowed sets with application to linguistic word modelling, Information Sciences, (2019), doi:10.1016/j.ins.2018.11.018.
[4] P.A. Gonzlez, J.M. Zamarreo, Prediction of hourly energy consumption in buildings based on a feedback artificial neural network, Energy and Buildings 37 (6) (2005) 595–601.
[5] Z. Yu, F. Haghighat, B.C.M. Fung, H. Yoshino, A decision tree method for building energy demand modeling, Energy and Buildings 42 (10) (2010) 1637–1646.
[6] F. Martnezlvarez, A. Troncoso, J.C. Riquelme, J.S. Aguilarruiz, Energy time series forecasting based on pattern sequence similarity, IEEE Transactions on Knowledge Data Engineering 23 (8) (2011) 1230–1243.
[7] X. Li, Y. Deng, L. Ding, L. Jiang, Building cooling load forecasting using fuzzy support vector machine and fuzzy c-mean clustering, in: International Conference on Computer and Communication Technologies in Agriculture Engineering(cctae), 2010, pp. 438–441.
[8] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, Energy and Buildings 37 (5) (2005) 545–553.
[9] Q. Li, Q. Meng, J. Cai, H. Yoshino, A. Mochida, Applying support vector machine to predict hourly cooling load in the building, Applied Energy 86 (10) (2009) 2249–2256.
[10] Y. Guo, J. Wang, H. Chen, G. Li, J. Liu, C. Xu, R. Huang, Y. Huang, Machine learning-based thermal response time ahead energy demand prediction for building heating systems, Applied Energy 221 (2018) 16–27.
[11] Z. Wang, Y. Wang, R. Zeng, R.S. Srinivasan, S. Ahrentzen, Random forest based hourly building energy prediction, Energy and Buildings 171 (2018) 11–15.
[12] K. Amasyali, N. El-Gohary, Building lighting energy consumption prediction for supporting energy data analytics, Procedia Engineering 145 (2016) 511–517.
[13] J. Wang, G. Li, H. Chen, J. Liu, Y. Guo, S. Sun, Y. Hu, Energy consumption prediction for water-source heat pump system using pattern recognition-based algorithms, Applied Thermal Engineering 136 (2018) 755–766.
[14] E. Mocanu, P.H. Nguyen, M. Gibescu, W.L. Kling, Deep learning for estimating building energy consumption, Sustainable Energy Grids and Networks 6 (2016) 91–99.
[15] G. Shi, D. Liu, Q. Wei, Energy consumption prediction of office buildings based on echo state networks, Neurocomputing 216 (2016) 478–488.
[16] C. Li, Z. Ding, D. Zhao, J. Yi, G. Zhang, C. Li, Z. Ding, D. Zhao, J. Yi, G. Zhang, Building energy consumption prediction: An extreme deep learning approach, Energies 10 (10) (2017) 1525.
[17] K. Chen, J. Jiang, F. Zheng, K. Chen, A novel data-driven approach for residential electricity consumption prediction based on ensemble learning, Energy 150 (2018) 49–60.
[18] G. Fu, Deep belief network based ensemble approach for cooling load forecasting of air-conditioning system, Energy 148 (2018) 269–282.
[19] C. Li, Z. Ding, J. Yi, Y. Lv, G. Zhang, Deep belief network based hybrid model for building energy consumption prediction, Energies 11 (1) (2018) 242.
[20] M.L. Chalal, M. Benachir, M. White, R. Shrahily, Energy planning and forecasting approaches for supporting physical improvement strategies in the building sector: A review, Renewable and Sustainable Energy Reviews 64 (2016) 761–776.
[21] S. Banihashemi, G. Ding, J. Wang, Developing a hybrid model of prediction and classification algorithms for building energy consumption, Energy Procedia 110 (2017) 371–376.
[22] B. Dong, Z. Li, S.M.M. Rahman, R. Vega, A hybrid model approach for forecasting future residential electricity consumption, Energy and Buildings 117 (2016) 341–351.
[23] Z. Wang, Y. Wang, R. Ssrinivasan, A novel ensemble learning approach to support building energy use prediction, Building Energy Efficiency 159 (2018) 109–122.
[24] L. Li, Y.-L. Lin, N.-N. Zheng, F.-Y. Wang, Parallel learning : a perspective and a framework, IEEE/CAA Journal of Automatica Sinica 4 (3) (2017a) 389–395.
[25] L. Li, Y.-L. Lin, D.-P. Cao, N.-N. Zheng, F.-Y. Wang, Parallel learning a new framework for machine learning, Acta Automatica Sinica 43 (1) (2017b) 1–8.
[26] Y. Lv, Y. Chen, L. Li, F.Y. Wang, Generative adversarial networks for parallel transportation systems, IEEE Intelligent Transportation Systems Magazine (2018).
[27] T. Liu, B. Tian, Y.F. Ai, L. Li, D.P. Cao, F.-Y. Wang, Parallel reinforcement learning: a framework and case study, IEEE/CAA Journal of Automatica Sinica 5 (4) (2018) 827–835.
[28] Y. Lin, X. Dai, L. Li, F. Wang, Pattern sensitive prediction of traffic flow based on generative adversarial framework, IEEE Transactions on Intelligent Transportation Systems (2018), doi:10.1109/TITS.2018.2857224.
[29] C. Huang, C. Moraga, A diffusion-neural-network for learning from small samples, International Journal of Approximate Reasoning 35 (2) (2003) 137–161.
[30] D.C. Li, C.S. Wu, T.I. Tsai, Y.S. Lina, Using mega-trend-diffusion and artificial samples in small data set learning for early flexible manufacturing system scheduling knowledge, Computers and Operations Research 34 (4) (2007) 966–982.
[31] D.C. Li, C.C. Chen, C.J. Chang, W.K. Lin, A tree-based-trend-diffusion prediction procedure for small sample sets in the early stages of manufacturing systems, Expert Systems with Applications 39 (1) (2012) 1575–1581.
[32] Z.S. Chen, B. Zhu, Y.L. He, L.A. Yu, A pso based virtual sample generation method for small sample sets: Applications to regression datasets, Engineering Applications of Artificial Intelligence 59 (2017) 236–243.
[33] P. Thanathamathee, C. Lursinsap, Handling imbalanced data sets with synthetic boundary data generation using bootstrap re-sampling and adaboost techniques, Pattern Recognition Letters 34 (12) (2013) 1339–1347.
[34] H.F. Gong, Z.S. Chen, Q.X. Zhu, Y.L. He, A monte carlo and pso based virtual sample generation method for enhancing the energy prediction and energy optimization on small data problem: An empirical study of petrochemical industries, Applied Energy 197 (2017).
[35] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: International Conference on Neural Information Processing Systems, 2014, pp. 2672–2680.
[36] K.-F. Wang, W.-M. Zuo, Y. Tan, T. Qin, L. Li, F.-Y. Wang, Generative adversarial networks from generating data to creating intelligence, Acta Automatica Sinica 44 (5) (2018) 769–774.
[37] N.-Z. Zheng, J.-L. Ding, Regression gan based prediction for physical properties of total hydrogen in crude oil, Acta Automatica Sinica 44 (5) (2018) 915–921.
[38] G. Douzas, F. Bacao, Effective data generation for imbalanced learning using conditional generative adversarial networks, Expert Systems with Applications 91 (2018) 464–471.
[39] Y. Li, Q. Pan, S. Wang, T. Yang, E. Cambria, A generative model for category text generation, Information Sciences 450 (2018) 301–315.
[40] C. Esteban, S.L. Hyland, G. Rtsch, Real-valued (medical) time series generation with recurrent conditional gans (2017) arXiv:1706.02633.
[41] B. Tang, Y. Tu, S. Zhang, Y. Lin, Digital signal modulation classification with data augmentation using generative adversarial nets in cognitive radio networks, IEEE Access PP(99) (2018) 1–1.
[42] J.F. Torres, A.M. Fernndez, A. Troncoso, F. Martnez-lvarez, Deep learning-based approach for time series forecasting with application to electricity load, in: International Work-Conference on the Interplay Between Natural and Artificial Computation, 2017, pp. 203–212.
[43] H.Y. Chen, D.C. Li, Generating information-diffusion-based virtual samples to improve small data set prediction for ceramic powder: A case study, in: International Conference on Applied Computing and Information Technology/ International Conference on Computational Science and Intelligence, 2015, pp. 374–378.
[44] J. Li, J.H. Cheng, J.Y. Shi, F. Huang, Brief introduction of back propagation (bp) neural network algorithm and its improvement, Advances in Computer Science and Information Engineering 169 (2012) 553–558.
[45] G.B. Huang, Q.Y. Zhu, C.K. Siew, Extreme learning machine: Theory and applications, Neurocomputing 70 (1) (2006) 489–501.
[46] A.J. Smola, B. Schlkopf, A tutorial on support vector regression, Statistics & Computing 14 (3) (2004) 199–222.
[47] M.M. Adankon, M. Cheriet, Support vector machine, Computer Science 1 (4) (2002) 1–28.