Journal of Physics: Conference Series

Prediction model of household appliance energy consumption based on machine learning

Prediction model of household appliance energy consumption based on machine learning

Lei Xiang[1], Tao Xie[1*] and Wei Xie[1]

[1]College of Computer, National University of Defense Technology, Changsha, Hunan, 410073, China

[*]Corresponding author's e-mail: hamishxie@vip.sina.com

Abstract. This paper presents and discusses the prediction model of household appliance energy consumption. The paper discusses data filtering to remove non-predictive parameters, and used for model training. Five prediction models are trained with repeated cross validation and evaluated in the testing set: support vector machine (SVM), k nearest neighbor (KNN), random forest (RF), extreme random forest (ERF), and long short-term memory network (LSTM). Among them, the LSTM, which is a deep learning method, behaves the best performance. LSTM has the highest R^2 (0.97) and the lowest RMSE (21.36) in the testing set. The analysis results show that the deep learning method has advantages in the prediction of household appliance energy consumption.

1. Introduction

With the continuous development of China's cities and the growth of residential construction, the energy consumption of residential buildings has increased in recent years. The increase in the economic level of residents has led to an increase in the purchase of household appliances. The survey shows that 57% of the energy consumption of residential buildings comes from household appliances, including computers, televisions, washing machines, refrigerators, rice cookers, electric water heaters and so on [1]. Relevant government departments have also attached great importance to the rapid growth of residential energy consumption, and carried out large-scale surveys of residential energy consumption. Researchers have studied between residential electricity consumption and per capita income, electricity prices, gas prices, and climatic conditions. The relationship predicts the development trend of domestic electricity consumption [2]. Forecasting residential energy consumption not only helps to understand the energy consumption of residential buildings in China, but also provides analysis data for China's energy conservation and emission reduction work, and can also be used to detect abnormal energy consumption patterns [3]. The electricity consumption of a house is mainly related to the type and quantity of household appliances, and the household appliances have an influence on the indoor environment, such as temperature, humidity, and light. Therefore, it is possible to predict the energy consumption of household appliances by establishing relevant models and using different environmental data and electricity consumption data [4].

For the prediction of household appliance energy consumption, the typical research in traditional machine learning methods is to use multiple regression, neural network, support vector machine and other methods to establish related models to predict the energy consumption of household appliances [5,6]. The model usually considers the time of day, outdoor temperature, rainfall index, wind speed, global solar radiation and other parameters. Long short-term memory is commonly used in deep learning methods, it has great advantages in time series simulation and often achieves good results [7].

In this paper, support vector machine, k-nearest neighbor, random forest, extremely random forest, long short-term memory are used to build models for predicting the energy consumption of household appliances, and the performance between them are compared.

2. Research method

2.1. Traditional machine learning method
Four traditional machine learning methods are used to build household appliance energy consumption prediction models: support vector machine (SVM), k nearest neighbor (KNN), random forest (RF), extreme random forest (ERF). SVM is a kind of generalized linear classifier for binary classification of data according to supervised learning. SVM was proposed in 1964, and it developed rapidly after the 1990s and derived a series of improved and extended algorithms. SVM can also be used to solve regression problems [8]. KNN can be used for regression by obtaining the nearest neighbors of a sample and assigning the average of the properties of these neighbors to the sample to obtain the properties of the sample. Another improved method is to give different weights to the influence of neighbors of different distances on the sample, such as the weight is inversely proportional to the distance. RF is an important Bagging-based integrated learning method that can be used for classification and regression. This paper uses a random forest regression model. The ERF algorithm is very similar to the RF algorithm. It is composed of many decision trees, introduces more randomization than random forests, and can solve the variance problem more efficiently. Its computational complexity is also slightly reduced.

2.2. Deep learning method
The deep learning method LSTM is used to build prediction model. LSTM is a time recurrent neural network suitable for processing and predicting important events with relatively long intervals and delays in time series. LSTM is a special type of recurrent neural network (RNN). Like other neural networks, it consists of an input layer, one or more hidden layers, and an output layer. The neurons in the hidden layer can not only receive from the input layer, but also receive information that neurons have perceived from the last moment [9]. This loop structure enables LSTM to learn the intrinsic characteristics of time series data and learn long-term dependency information as show in figure 1 and figure 2. LSTM was introduced in 1997 and has been improved and promoted in the near future, due to its unique design structure, LSTM is suitable for processing and predicting important events with very long intervals and delays in time series

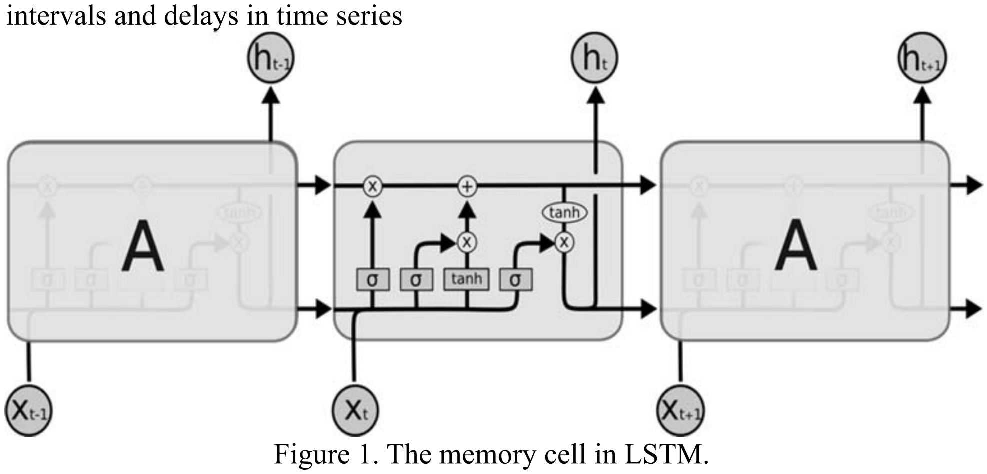

Figure 1. The memory cell in LSTM.

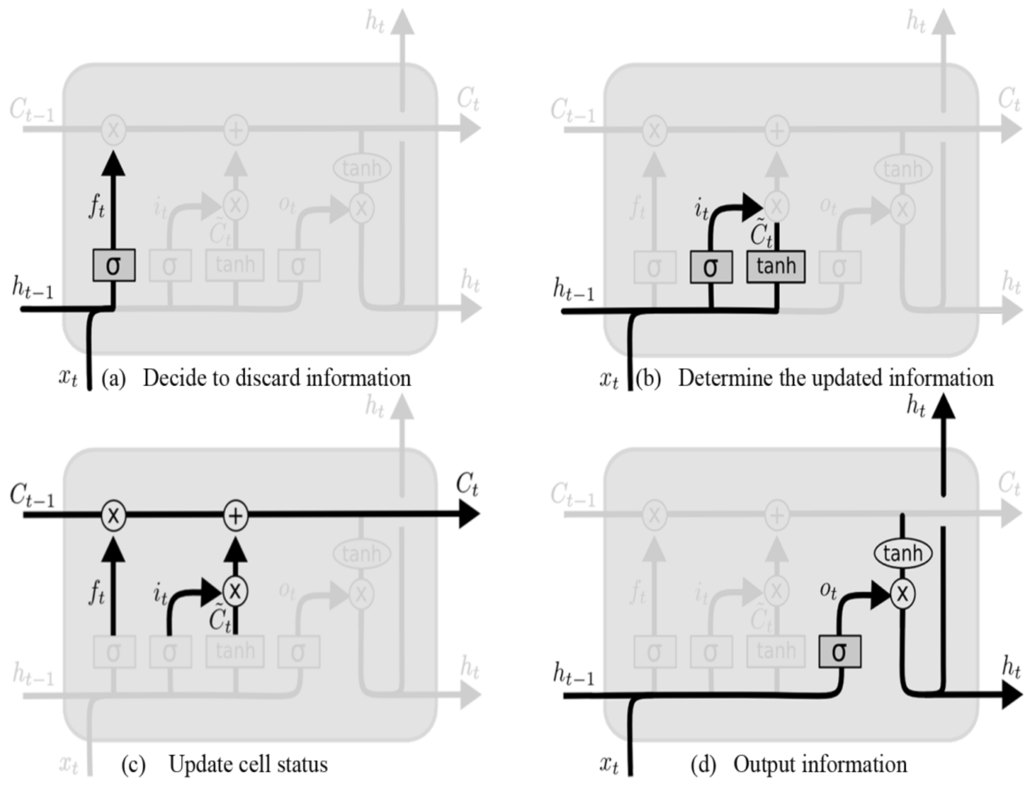

Figure 2. Internal structure of the LSTM memory unit.

The specific process is as follows:

The first step in LSTM is to decide what information we will discard from the cell state. This decision is made through a process called the forgotten gate. The gate reads h_{t-1} and x_t and outputs a value between 0 and 1 for each number in cell state C_{t-1}. 0 means completely reserved and 1 means completely discarded. As shown in (a) of figure 2, the calculation formula of the forgetting gate f_t is as follows:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \tag{1}$$

The next step is what new information will be stored in the cell state. This is called the input gate and determines what value we want to update. Then a tanh layer creates a new candidate value vector $\tilde{C}_t$ and adds it to the state. As shown in (b) of figure 2, the calculation formula for the information update is as follows:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \tag{2}$$

$$\tilde{C}_t = tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \tag{3}$$

As shown in (c) of figure 2, the previous steps have already decided what will be done, then update the neuron status with the results obtained earlier, equation as follows:

$$\tilde{C}_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{4}$$

Finally, we need to determine what value to output as shown in (d) of figure 2, that is, calculate the output gate O_t, O_t can determine how many hidden layer state variables h_t at time t, as show in equation (5) and equation (6):

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \tag{5}$$

$$h_t = o_t * tanh(C_t) \tag{6}$$

where x_t is the input at time t, W_i, W_C, W_f, W_o are weight matrices, b_f, b_i, b_C, b_o are biases, C_t, $\tilde{C}_t$ are the new state and candidate state of memory cell, f_t, o_t are forget gate and output gate.

3. Experimental results and discussion

3.1. Data sets

3.1.1. Data sets description
The data for this experiment is from public data on GitHub, which was obtained by continuously measuring a low-energy house in Belgium for 137 days, including electrical energy consumption data and environmental data. Energy consumption data is collected every 10 minutes to a house inside the electrical power consumption, environmental data is the use of wireless sensor network monitoring temperature and humidity, including living room, kitchen, bathroom, laundry and so on. The sensor collects data every ten minutes. The data sets also joined the nearby airport meteorological data, and through the date and time and electric energy consumption data is merged. The data sets contains 19735 sets of data, including 27 features, as shown in table 1.

Table 1. Data variables and description.

Number of features	Data variables	Description	Units
1	T1	Temperature in kitchen	°C
2	T2	Temperature in living room	°C
3	T3	Temperature in laundry room	°C
4	T4	Temperature in office room	°C
5	T5	Temperature in bathroom	°C
6	T6	Temperature outside the building	°C
7	T7	Temperature in ironing room	°C
8	T8	Temperature in teenager room	°C
9	T9	Temperature in parents room	°C
10	R1	Humidity in kitchen	%
11	R2	Humidity in living room	%
12	R3	Humidity in laundry room	%
13	R4	Humidity in office room	%
14	R5	Humidity in bathroom	%
15	R6	Humidity outside the building	%
16	R7	Humidity in ironing room	%
17	R8	Humidity in teenager room	%
18	R9	Humidity in parents room	%
19	L	Light energy consumption	Wh
20	RO	Humidity outside the airport	%
21	Td	Dew point temperature	°C
22	V	Visibility	km
23	W	Wind speed	m/s
24	TO	Temperature outside the airport	°C
25	rv1	Random Variable1	\
26	rv2	Random Variable2	\
27	P	Pressure	mmHG

Figure 3 shows the energy consumption graph of the appliance during the whole measurement period and the energy consumption graph of the first week. It can be seen that the energy consumption data changes greatly in the time series.

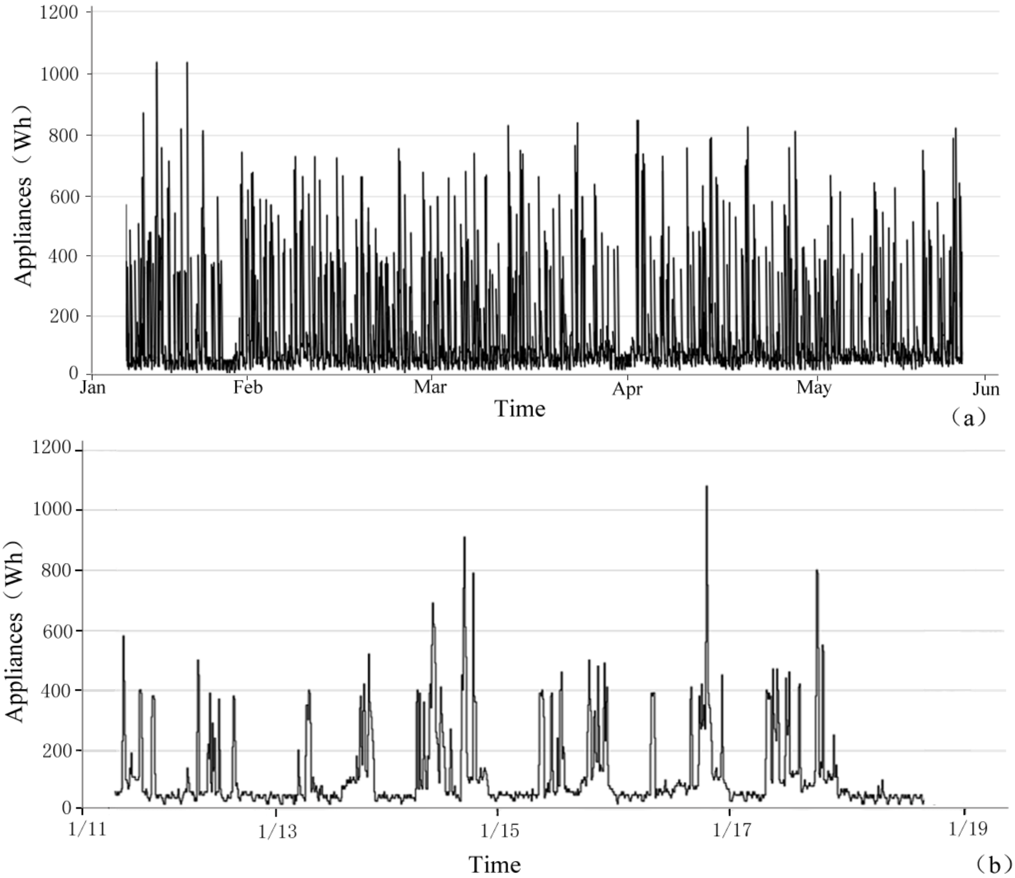

Figure 3. (a) Household appliance energy consumption in the whole period,
(b) Household appliance energy consumption in the first week.

3.1.2. Data features selection

In the training of machine learning models, if there are too many useless features, the accuracy of the model will be reduced. Before the model training, feature selection can reduce over-fitting, improve accuracy and shorten model training time. There are usually three methods for feature selection. The filtering method: to score each feature according to divergence or correlation, to set the threshold or the number of thresholds to be selected, and to select features. The feature screening method: by continuously excluding features or continuously selecting features, and the effect of the model obtained by training is scored, and the feature retention is determined by predicting the effect score. The embedding method: first to use some machine learning algorithms and models to train, then to obtain the weight coefficient of each feature, selecting features. according to the coefficient from large to small, it's similar to the filter method, but the former is through training to determine the pros and cons of features [10]. In this experiment, the filtering method is used to measure the correlation between variables by the pearson correlation coefficient (r). The formula is as follows:

$$r = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \overline{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}} \tag{7}$$

The range of r is between [-1,1], and the closer |r| is to 1, the higher the correlation between variables, and the closer to 0, the lower the correlation between variables. Figure 4 is the correlation between all the characteristic variables and between the characteristic variables and the energy consumption. It can be seen from the figure that all temperature characteristics from T1~T9 are positively correlated with energy consumption. For indoor temperature, the correlation coefficient between T9 and T3, T5 and T7 is greater than 0.9, which is highly positively correlated with them. For outdoor temperatures, T6 has a correlation with TO of 0.97 and is highly positively correlated. Analysis of the two temperature characteristics of T9 and T6, the information they provide can be provided by other temperature data, so it can be removed from the data sets. The correlation coefficient between wind speed V and energy consumption in environmental data is 0. It can be said that there is no correlation. The correlation coefficient between two random variables, rv1 and rv2, is 0.01, and the correlation is extremely low, so it has little effect. V and two random variables are removed from the data sets.

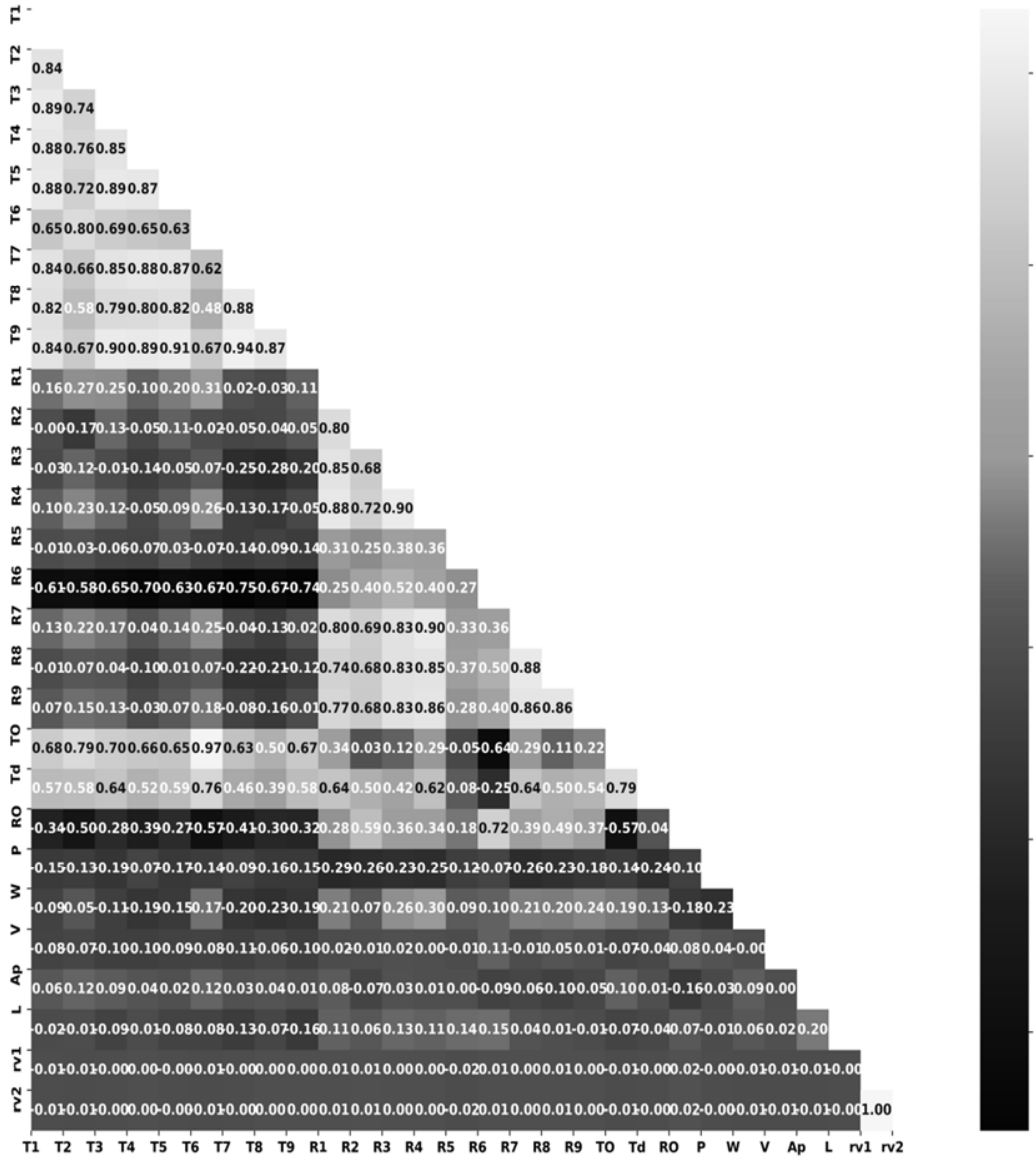

Figure 4. Correlation coefficient between characteristic variables.

3.1.3. Data sets partition
After feature selection, T9, T6, r1, r2, and V in the data sets are eliminated, and the number of features in the data sets is reduced to 22, but the feature quantity is still large, and the eigenvalue unit and the order of magnitude have large differences, so feature scaling is performed to normalize the data and

scale all feature data values between [0, 1]. As shown in equation (8):

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (8)$$

Cross-validation is used in this experiment, so the data sets is randomly divided into training sets and testing sets, 75% of which is divided into training sets for training model, and the remaining 25% is used as testing sets for testing model. As shown in table 2, the data sets are divided.

Table 2. Training Sets and Testing Sets.

Data sets	Group of data sets	Number of variables
Training sets	14801	22
Testing sets	4934	22

3.2. Model evaluation indicators

The prediction models established in the experiment are all regression models. The commonly used indicators for evaluating regression models are root mean square error (RMSE), mean absolute error (MAE), mean square error (MSE), and decision coefficient (R^2). The following two evaluation indicators, RMSE and R^2, are used in this experiment. RMSE is the square root of the ratio of the squared deviation of the observed value to the true value and the ratio of the observed number. It is used to measure the deviation between the observed value and the true value. R^2 is the square value of R, which reflects that all the variation of the dependent variable can be the proportion of the independent variable interpretation. In the regression model, the smaller the evaluation index RMSE, the larger the R^2, indicating that the model is better. The calculation formula of RMSE and R^2 are shown in equation (9) and equation (10).

$$RMSE = \sqrt{\frac{1}{\mathrm{m}} \sum_{\mathrm{i}=1}^{\mathrm{m}} (y_i - \hat{y}_i)} \quad (9)$$

$$R^2 = 1 - \frac{\sum_{i=1}^{m} (\hat{y}_i - y_i)^2}{\sum_{i=1}^{m} (y_i - \overline{y}_i)^2} \quad (10)$$

Where y_i is the real value of the data at the time points i, $\hat{y}_i$ is the predicted value of the data at the time points i, and $\overline{y}$ is the average value of the sample, m is the total number of samples.

3.3. Model parameters

In the experiment, the prediction model based on the traditional machine learning method has many parameters, so the grid search method is used to determine the optimal parameters of the model. Grid search is an exhaustive search method for specifying parameter values. The parameters of the estimation function are optimized by cross-validation method to obtain optimal parameters. The possible values of each parameter are arranged and combined to list all possible parameters. Combine the results to generate a "grid", then use each combination for training in the model, and use cross-validation to evaluate the performance and automatically adjust to the optimal combination of parameters based on the selected scoring method [11]. The experiment uses R^2 and RMSE as the optimization indicators, and has undergone ten cross-validation. All experiments are done in python.

After parameter optimization, the parameters of KNN model are chosen as algorithm (brute), neighbors (3), weights (uniform), and the parameters of the SVM model are C (8), gamma(1), and kernel (rbf). The RF model parameters are selected as max depth (40), max features (sqrt), estimators

(240), and ERF model parameters are selected as max depth (80), max features (sqrt), estimators (250).

In the LSTM prediction model, there are 50 neurons in the hidden layer and 1 neuron in the output layer. The model uses 500 epochs, and each batch has a size of 10. Adam is used for model optimization algorithm, and model selects RMSE as the objective function in the LSTM optimization process.

3.4. Discussion of results
Table 3 shows the performance evaluation results of five household appliance energy consumption prediction models. According to the evaluation index, the model with lower RMSE and higher R^2 will have better performance. It can be seen from the table that ERF has the highest R^2 in the training set, reaching 0.99, and LSTM has the smallest RMSE, only 2.48. LSTM performs best in the training set, with the worst performing SVM, R^2 being the lowest 0.83 and RMSE being the highest 41.13. In the testing set, LSTM has the highest R^2 (0.97) and lowest RMSE (21.36), which performs best. Among models based on traditional machine learning methods, KNN, SVM, and RF have similar performance in the testing set but they are not as good as ERF. In general, the generalization of the four traditional machine learning methods is not particularly good, they perform well in the training set, but not well in the testing set. The best method is LSTM, which performs well in both the training set and testing set.

Table 3. Model performance.

Model	Training		Testing	
	R^2	RMSE	R^2	RMSE
KNN	0.86	**4. 37.08**	0.58	64.99
SVM	0.83	41.13	0.57	65.52
RF	0.94	23.84	0.57	65.64
ERF	0.99	9.99	0.64	59.81
LSTM	0.97	2.48	0.97	21.36

Figure 5 shows the comparison of the real and predicted values of the LSTM prediction model on the household appliance energy consumption data sets. It can be seen that the predicted value of the model is very close to the real value, LSTM prediction model performs very well.

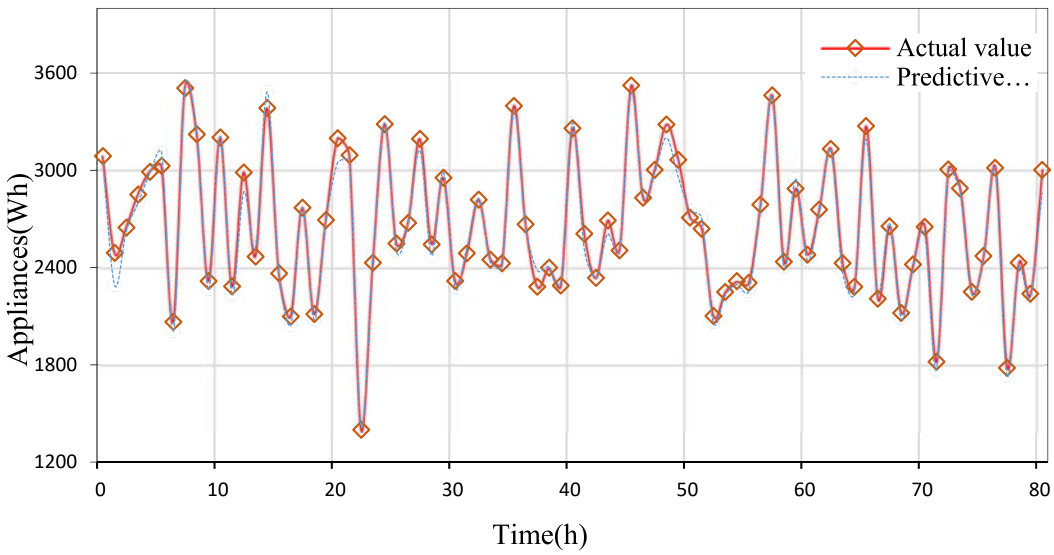

Figure 5. Prediction results of household appliance energy consumption data by LSTM model.

4. Conclusion

The household appliance energy consumption prediction models based on SVM, KNN, RF, ERF, and LSTM are explored. First, we study the preprocessing of data, remove some features in the filtered data and normalize the data. Second, the grid search method is used to determine the optimal parameters in the model, and the models based on different machine learning methods are established. Finally, the prediction performance of each model was evaluated and compared. The results show that among the four prediction models established by the traditional machine learning method, ERF can achieve good results in the training set and the testing set, with the best prediction performance. KNN, RF and SVM have similar prediction performance in the testing set, but SVM has the worst performance in the training set. The four models based on traditional machine learning are all relatively average in predicting performance. The prediction performance of the model based on LSTM is much better than that of the model based on traditional machine learning, and the predicted value of LSTM model is very close to the real value, it has advantages in terms of household appliance energy consumption prediction.

References

[1] Xiao, X.R., Hu, S., Yan, D. (2012) Research on Electric Model of Household Appliances Based on Actual Measurement. Building science, 28:223–231.
[2] Jian, H.Y., Kong, W.J., Liu, H.X. (2013) Study on Influences of Lighting and Appliances Usage Behavioral Patterns on Household Electricity Consumption. Building science, 29:15–19.
[3] Chen, S.Q., Li, N.P., Fu, X.Z. (2007) Statistical methods of energy consumption for residential buildings. Heating Ventilation &Air Conditioning, 37:44–48.
[4] Zhou, J., Song, Y.L. (2018) Study on the residential Thermal Environment and Energy Consumption Based on Human Occupancy. Building science, 34:59–65.
[5] Candanedo, L.M., Feldheim, V, Deramaix D. (2017) Data driven prediction models of energy use of appliances in a low-energy house. Energy and Buildings, 140:81–97.
[6] Tsanas, A., Xifara A. (2012) Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools. Energy Build, 49:560–567.
[7] Sha, G.R., Qian, Q. (2018) Application Reviews of Artificial Intelligence Algorithm in Building Energy Consumption Prediction. Sci-Techinnovation & Productivity, 296:58–62.
[8] Kwok, T.Y. (2000) Support vector mixture for classification and regression problems. In: Proc of 14th International Conference on Pattern Recognition, Brisbane. pp. 255–258.
[9] Yang L., Wu Y.Q., Wang J.L. (2018) Research on recurrent neural network. Journal of Computer Applications, 38:6-11.
[10] Cetin, K.S., Tabares-Velasco, P.C., Novoselac, A. (2014) Appliance daily energy use in new residential buildings: Use profiles and variation in time-of-use. Energy Build, 84:716–726.
[11] Wang J.F, Zhang L., Chen G.X. (2012) A parameter optimization method for an SVM based on improved grid search algorithm. Applied Science and Technology, 39:28-31.