

Regional energy consumption prediction based on SARIMAX-LSTM model

Yang Zhou^{1,*}

¹*School of Control and Computer Engineering, North China Electric Power University, Beijing 102206, China*

*Corresponding author: 791069624@qq.com

Abstract: *Accurate prediction of energy consumption is helpful for further value mining and data fusion. In order to achieve this purpose, this paper proposes an energy consumption prediction method based on SARIMAX (seasonal autoregressive integrated moving average with exogenous) and LSTM (Long Short-Term Memory) hybrid model. First, we import energy consumption data and weather environment factors such as temperature, moderateness, wind speed, etc., compare the corresponding relationships through data visualization, and use K-means to construct weather clusters as an exogenous variable. Then we import holiday information, and construct a holiday indicator as the second exogenous variables. Furtherly, we adjust the model according to the seasonal trend and use grid search to select the optimal parameters of the SARIMAX model. Finally, we mix the SARIMAX model with LSTM to optimize the prediction model to make predictions and compare the results. Experimental analysis shows that the SARIMAX-LSTM hybrid model can integrate weather factors, holiday factors, and seasonal factors to make high-precision predictions of energy consumption.*

Keywords: SARIMAX, LSTM, K-means, prediction of energy consumption

1. Introduction

Our country is a big energy consuming country. Different regions have different levels of energy consumption according to their population density, development level, geographic location and other factors. Accurate energy consumption predictions for specific regions can make the most of energy and reform energy utilization mode. As domestic efforts have been made to build ubiquitous power Internet of Things [1] and urban energy Internet [2], high-precision energy consumption prediction results can provide decision-makers with judgment basis and reference value to help building smart grid [3] and smart city [4]. Energy consumption data is mostly stored in a structured form, which belongs to time series data. Data itself will be affected by some external factors (weather, holidays, seasonal alternations, etc.), so that the data sequence itself will have a greater impact on specific nodes. The data fluctuate which has a random tendency makes predictions more difficult.

The time series prediction can be traced back to the ancient Egyptians' long-term observation and prediction of the fluctuations of the Nile River 7000 years ago, and then the time series prediction work has been researched throughout the history of mankind. With the development of modern computer technology and the application of statistical methods, the British scientist George Udny Yule created a stationary linear autoregressive model for the research object of time series, which pioneered the first time domain analysis of time series. The AR (auto regressive) model proposed by Yule believes that there is a correlation between time series variables, and the time series is not a function of time. Then Gilbert Thomas Walker added correlation coefficients to Yule's AR model to further reduce the impact of accidental factors. Afterwards, Russian probability theorists Evgenievich Slutsky focused on the random component nature of time series data and proposed a MA (moving average) model based on error perturbation. With the further development of AR and MA model theory, Sweden's famous econometrics Herman Word took discrete stationary random processes as his research object, and strictly proved that discrete stationary processes are composed of hidden periods and linear regression processes. Once the hidden periods are eliminated as deterministic components, model disturbances can be effectively reduced. The linear regression process is caused by sliding the composition of averaging and autoregressive process. This decomposition idea contributed to the generation and development of the ARMA model fitting stationary series, and then along with George Edward Pelham Box systematically proposed the ARIMA (auto regressive integrated moving average) model which has complete research

methods including modeling, analysis, and forecasting, the theory and practice of time series forecasting have developed vigorously. Nowadays, the development of time series forecasting has gradually been combined with machine learning methods, neural networks and deep learning networks which can deeply mine the potential value of data and further accurately grasp the future trend of time series data. This article mainly studies the integration of traditional time series prediction models and machine learning models on this basis to achieve better prediction results.

There are roughly four forecasting methods based on time series which are linear regression using time characteristics, traditional linear time series model modeling, deep learning network construction, and transformation of time series data sets into supervised learning data sets and remodeling. 1) Linear regression methods using time characteristics usually express time as 0-1 dummy variables, such as seven one-hot codes from Monday to Sunday, thereby eliminating the dependence of time series data and reducing the excessive influence of historical data on the prediction results. Fan Lijun [5] applied it to energy consumption prediction and achieved good results. This method can better remove the dependence of the data itself based on historical information, but it only considers the time series data itself without the external information that may affect the data; 2) Traditional ARMA/ARIMA and other time series models use d-differentiation to smooth the time series, and perform white noise test on the residual square. This method mainly focuses on the autocorrelation of the series itself in order to reduce the impact of errors. Lu Chengju [6] applied it to the energy consumption prediction of diesel passenger cars. After removing the trend term and the second-order difference, the originally unstable diesel energy consumption time series data was smoothed and the ARIMA model was constructed. Finally the model obtained a high-precision prediction results. This method can better deal with the impact of white noise on the collection results during the data collection process, but it fails to take the possible seasonal trend effects of the data itself into account; 3) The deep learning network mainly uses CNN to capture short-term local dependencies and uses RNN to capture long-term macro dependencies. Guokun Lai [7] proposed LSTNet based on this structure to solve the mixed situation of short-term and long-term models involved in time series data prediction. This method has better prediction accuracy for time series data with a mixture of long short-term models, but a single deep learning network will take a lot of time to complete the training process of the entire model and may cause overfitting; 4) Through the functions under the pandas framework in the python environment, the time series data set can be transformed to the supervised learning data set, and then the commonly used supervised learning models (xgboost, LSTM, TCN, seq2seq, etc.) can be used to achieve prediction function. Ji Zengge [8] utilized this method on the basic of the mutual information entropy (Mutual Information Entropy, MIE) correlation measurement index and finally established the mapping relationship between meteorological factors and photovoltaic power through the LSTM model. This method transforms the prediction problem of time series data into a more diversified way of solving the prediction problem of supervised learning data sets, and provides more possibilities for the selection and optimization of prediction models. However, this method requires the dimensionality of the original data to be improved. When facing with a huge amount of data, it will exponentially increase computing and storage costs.

The above four types of methods all have their own advantages and disadvantages, and they have different levels of application in different scenarios. However, real application scenarios are usually changeable. This requires energy consumption prediction model to take various factors into account to achieve maximum universal use. On this basis, this paper proposes a time series energy consumption prediction method based on the SARIMAX-LSTM hybrid model. First, weather information and holiday information are introduced as external reference information, and then external information is used as exogenous variables together with seasonal differences to establish a SARIMAX model. Afterwards, our method convert the output of the SARIMAX model into the supervised learning data set, and input the converted data set into LSTM model which is good at processing a mixture of short-term and long-term models .It can help making predictions maximize the value of historical information and reduce the impact of errors caused by useless historical information. At this time, because the training of the original data has been completed by the SARIMAX model, the computational cost of the data set conversion process and the training and prediction process with the LSTM model will be significantly reduced while maintaining good prediction accuracy. The method proposed in this paper comprehensively considers external factors, seasonal trends, historical information, and the computational cost of supervised learning methods. It is suitable for a variety of energy consumption prediction scenarios.

2. Construction of SARIMAX-LSTM hybrid model

Traditional time series forecasting tasks are mostly based on a single model. Hyperparameters are adjusted for specific forecast data sets to achieve the forecast results. However, a single model often has

its weaknesses. When faced with suitable applicable scenarios or self-generated data It can better reflect the performance of its algorithm, but it is difficult to guarantee the applicability and prediction performance of the model when faced with actual complex scene sampling data that is affected by many factors during sample collection. In response to the above-mentioned problem, this paper selects a hybrid model composed of a SARIMA model that can add additional influencing factors and consider seasonal trends separately and an LSTM model that can effectively deal with a mixture of long and short-term models to complete energy consumption prediction tasks. The hybrid model integrates weather, holidays, influence of seasonal trends and the value of long-distance historical information. Our model also uses adaptive parameter select methods so that can significantly improve prediction effect.

2.1 Data preparation

The time series data only has two columns of variables which are time and data at the time point, so there is no problem of inconsistent dimensions. In the preprocessing work, only simple null and redundant values need to be handled. The main work of the data preparation stage is to complete the processing of exogenous variables, including the import of weather data and holiday information.

The weather data contains 31 columns of variables such as temperature, humidity, visibility, UV index, etc. In this article, we first try to reduce the number of parameters before doing subsequent processing using methods such as: (1) Only the average temperature is selected among the highest temperature, the lowest temperature, and the average temperature. The remaining two parameters are discarded; (2) The parameters with extreme low correlation, such as sunrise and sunset time, are discarded. These methods can minimize the number of parameters and reduce the burden of subsequent processing calculations and the impact of errors.

After selecting the parameters that need to be processed, this article will take time-weather factors as the processing object, and use the K-means algorithm [9] to cluster the weather conditions of each day. The K value determination method uses the Elbow Method [10]. Its implementation is to draw the cost function curve when k takes different values. The point where the slope of the curve changes the most is similar to a human elbow joint, which is the best value of K. Assuming that the data set to be processed is $X=[x^{(1)}, x^{(2)}, x^{(3)} \dots x^{(m)}]$, k clusters are $C=C_1, C_2, C_3 \dots C_k$, the goal is to minimize the loss function:

$$E = \sum_{i=1}^k \sum_{x \in C_i} \|x - \mu_i\|^2 \quad (1)$$

Where μ_i represents the cluster center of cluster C_i :

$$\mu_i = \frac{1}{|C_i|} \sum_{x \in C_i} x \quad (2)$$

K-means uses a greedy algorithm to avoid traversing all possible cluster divisions by finding approximate solutions. The specific steps are as follows:

- (1) Each cluster randomly selects a cluster center point to obtain a set of k center points $\{\mu_1, \mu_2, \mu_3, \dots, \mu_k\}$
- (2) Calculate the Euclidean distance $\text{dist}(x^{(l)}, \mu_j)$ from all sample points to each center point, and classify the sample points into the nearest class.
- (3) Recalculate each cluster center μ_i according to the newly divided clusters.
- (4) Repeat (2) and (3) until the change range of each cluster center is within the preset threshold or reaches the upper limit of iteration.

The holiday data is similar to the time series data. It has a two-column variable of time and holiday name, which does not require excessive processing means. The processing method selected in this article is to input the information as a 0-1 binary variable, 0 means that the day is not a holiday and 1 means the opposite.

2.2 Stationarity test and smoothing treatment

Stationarity test is performed by ADF (augmented Dickey-Fuller) test [11]. The test method is to determine whether the time series has a unit root. If it does, the sequence belongs to a non-stationary random walk sequence. If it doesn't, the sequence is considered to be stable.

The so-called random walk, consider the time sequence $\{y_t\}$, if the sequence satisfies

$$y_t = y_{t-1} + \epsilon_t \tag{3}$$

Then call the time series $\{y_t\}$ a random walk sequence, where ϵ_t is white noise, obeys $D(0, \sigma^2)$, and the random walk sequence with offset and trend terms is as follows:

$$y_t = y_{t-1} + \epsilon_t + \mu + \beta t \tag{4}$$

Consider the simple first-order autoregressive formula:

$$y_t = \gamma y_{t-1} + \epsilon_t \tag{5}$$

Subtracting y_{t-1} from both sides of equation (5), we get:

$$y_t - y_{t-1} = (\gamma - 1)y_{t-1} + \epsilon_t \tag{6}$$

Let $\Delta y_t = y_t - y_{t-1}$, $\delta = (\gamma - 1)$, we get:

$$\Delta y_t = \delta y_{t-1} + \epsilon_t \tag{7}$$

Let hypothesis $H_0: \delta = 0, H_1: \delta \neq 0$, when the null hypothesis H_0 is true, the corresponding sequence is considered to be a random walk sequence, which is non-stationary. Test by testing the t statistic:

$$t_\gamma = \frac{\hat{\gamma}-1}{s(\hat{\gamma})} \text{ or } t_\delta = \frac{\hat{\delta}}{s(\hat{\delta})} \tag{8}$$

Among them, $s(\hat{\gamma})$ and $s(\hat{\delta})$ are the standard deviations of the estimated values $\hat{\gamma}$ and $\hat{\delta}$, respectively.

Consider the following three test models:

$$\left\{ \begin{array}{l} \text{model1: } \Delta y_t = \delta y_{t-1} + \sum_{i=1}^p \lambda_i y_{t-i} + \epsilon_t \\ \text{model2: } \Delta y_t = \delta y_{t-1} + \sum_{i=1}^p \lambda_i y_{t-i} + \epsilon_t + \mu \\ \text{model3: } \Delta y_t = \delta y_{t-1} + \sum_{i=1}^p \lambda_i y_{t-i} + \epsilon_t + \mu + \beta t \end{array} \right. \tag{9}$$

The test usually start from the third one and then judge the necessity of fitting the second one and the first one according to whether β and μ are zero. We query the critical value table of τ under the given significance level condition. If $t_\gamma < \tau$, reject H_0 , and the sequence is considered to be stationary, and vice versa. In actual engineering, the P value, which is the probability value corresponding to the t statistic, is usually considered for judgment. Usually, when the P value is less than 0.05, the sequence is considered to be stable.

When the sequence is not stationary, the k-order difference is used to stabilize the sequence. The specific method is to subtract two sets of data with a mutual distance of k. The result is called the difference sequence. The difference sequence is generated from k=1. Use the ADF test results until a stable difference sequence is obtained, and the k-th generated difference sequence will lose k sets of data. The obtained stationary difference sequence is used as the model input, and the output prediction result is the difference value. At that time, only the difference value needs to be added to the current value to obtain the prediction result. When the sequence smoothing process is completed, the SARIMAX model can be constructed.

2.3 Building the SARIMAX model

When the sequence stability is guaranteed, we can start to estimate the parameters of the SARIMAX model [12] and evaluate each parameter combination. The main parameters that need to be determined are the autoregressive order p, moving average q, difference order d, the seasonally stable autoregressive order P, the seasonally stable moving average Q, the seasonal difference order D, and the seasonal period S. The seasonal period S is usually 12, because the study of seasonal trends usually takes one month as the minimum unit. The SARIMA model is shown as formula (10).

$$\gamma(B)\gamma(B^S)(1-B)^d(1-B^S)^D y_t = C + \theta(B)\theta(B^S)\epsilon_t \tag{10}$$

Where $\gamma(B)$ is the p-order autoregressive coefficient polynomial of the model, and $\theta(B)$ is the q-order moving average coefficient polynomial of the model, B is the back shift factor, which satisfies $B^n y_t = y_{t-n}$, B^S is the seasonal shift factor, $\varphi(B^S) = 1 - \varphi_1(B^S) - \dots - \varphi_p(B^{S,p})$ is the polynomial of the autoregressive coefficient of the seasonally stationary model, $\theta(B^S) = 1 - \theta_1(B) - \dots - \theta_q(B^{S,q})$ is the polynomial of the moving average coefficient of the seasonally stationary model.

There are usually two options for the optimal selection of model parameters:

(1) Using AIC Information Guidelines

AIC (akaike information criterion) [13] is a commonly used standard to measure the pros and cons of statistical models. AIC uses the maximum likelihood function to determine the pros and cons of the model training results. However, in actual engineering problems, not only the maximization needs to be considered Likelihood function, it also needs to consider the influence of model complexity. Therefore, when the AIC value is used as the judgment basis, the BIC (bayesian information criterion) value will also be considered. The formulas of AIC and BIC are as follows:

$$AIC = 2k - 2\ln(L) \tag{11}$$

$$BIC = k\ln(n) - 2\ln(L) \tag{12}$$

Where k is the number of parameters, L is the model likelihood function, and n is the number of samples. For AIC, when k increases, the complexity of the model will increase. At the same time, the likelihood function L will increase and the growth rate is greater than K, so that AIC becomes smaller. But when k is too large, the second term on the right side of the formula increases slowly, which leads to an increase in AIC. Therefore, selecting the parameter combination with the smallest AIC can improve the model fit while reducing the number of parameters k as much as possible, thereby avoiding the risk of overfitting. The $k\ln(n)$ term in BIC is used as a penalty term when there are too many parameters, and its penalty is significantly greater than 2k. At the same time, the model complexity caused by overfitting when the number of samples is too large is also considered.

(2) Using accuracy rate

For example, we can use accuracy index like MAE (Mean Absolute Error, Mean Absolute Error), MAPE (Mean Absolute Percentage Error, Mean Absolute Percentage Error) through the degree of error between the predicted value of the model and the actual value to judge the quality of the parameter combination. Assuming that the predicted sequence $\hat{Y} = \{\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n\}$ and the actual sequence $Y = \{y_1, y_2, \dots, y_n\}$, the MAE and MAPE calculation formulas are as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i| \tag{13}$$

$$MAPE = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{\hat{y}_i - y_i}{y_i} \right| \tag{14}$$

In this paper, AIC information criterion is used for parameter estimation, and the accuracy index is used to test the estimation result, comprehensively considering the actual data situation and the prediction target, and selecting the most suitable parameter combination.

2.4 Combining with the LSTM model

The LSTM network is a supervised learning model which requires both input value X and output value Y. The predicted value of Y is obtained through the linear combination of X, and the time series data is only composed of time points and time series data, and there is no input-output mapping. Therefore, it is necessary to transform the time series data set into a supervised learning data set. Considering that the value of time series data at the current time point is usually affected by historical data, this article adopts a feasible conversion method: use the fitted time series data obtained after the SARIMAX model training as the input value Y, copy the time series data and move it down as a new line, the first row of the new variable sequence is empty, use the new variable sequence as a column in the input variable X, continue to move the new sequence down in the same way, repeat this step, and finally get k input variables, remove the records with null values, horizontal comparison can be obtained for a certain row of data, the output value is the original fitted time series variable y_i , and the input value is the first k historical data of y_i . At the same time, the exogenous variable holiday information and

weather information conform to the definition of the input variable X in the supervised learning data set, and can also be input to increase the degree of model fitting. Among them, h_k represents whether the k-th group of data time is a holiday, and w_k represents the k-th group of data weather cluster number.

Table 1: Converting to supervised learning data set

var1(t-k)	...	var1(t-2)	var1(t-1)	Holiday	Weather	var1(t)
y_1	...	y_{k-1}	y_k	h_{k+1}	w_{k+1}	y_{k+1}
y_2	...	y_k	y_{k+1}	h_{k+2}	w_{k+2}	y_{k+2}
y_3	...	y_{k+1}	y_{k+2}	h_{k+3}	w_{k+3}	y_{k+3}
...

Through the above process, k+2 input variables and one output variable can be obtained. The structure meets the required structure of the supervised learning data set. Aiming at the dimensional inconsistency between the first k historical time series data and the latter two exogenous variables, this paper adopts the regularization operation to unify the magnitude. At this time, the new data set is imported into the LSTM model for fitting, combined with the trained model to make predictions, and the prediction results are compared with the actual data to prove the feasibility of the proposed scheme.

3. Experiment and evaluation

The experimental operating environment is an online cloud computing platform dedicated to the kaggle competition, the experimental tool is Jupyter Notebook which is a commonly used python tool, and the experimental data is the kaggle public data set Smart meters in London, including the total energy consumption of the 5567 London households in the Carbon London project and the weather data in London during February 2011 to November 2014. The sampling period is 1 day. The experiment divides the sample data into training set and test set according to 7:3 for model calibration, and finally predicts 30 sets of data to test the prediction effect. Model calibration uses MAE (Mean Absolute Error) and MAPE (Mean Absolute Percentage Error) as the calibration point, and the prediction effect is verified by drawing a comparison curve.

3.1 Creating weather cluster

Weather factors on the day of energy use, including temperature, humidity, wind speed, etc., will have varying degrees of impact on energy consumption. The method and extent of the impact will be determined according to the actual conditions of the specific area. This article will analyze the impact of different weather factors on energy consumption based on the comparison between energy consumption data and the local weather data collected together with it, and use the result as a basis to classify energy consumption data through K-means algorithm to classify weather clusters. The class identification is integrated into the energy consumption data set, and is ready to be used as the input value of subsequent models.

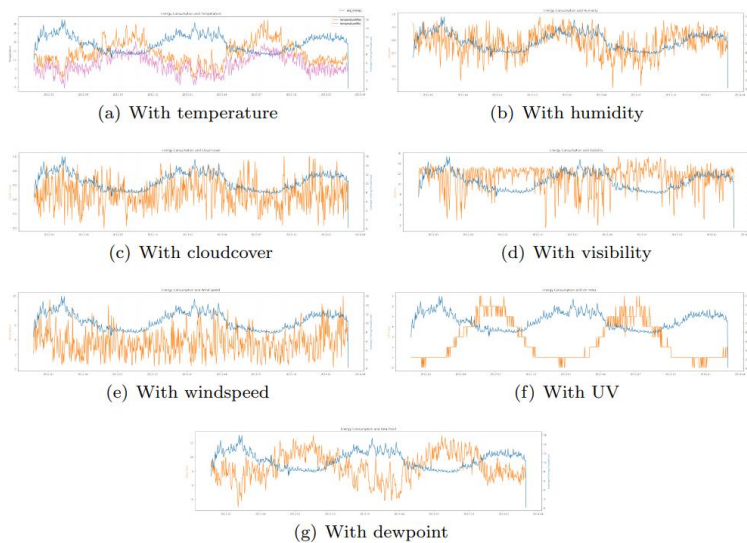


Figure 1: Comparison of the relationship between weather factors and energy consumption

The above relationship comparison chart is obtained by standardizing energy consumption data and weather data and adjusting the y-axis parameters of the plot drawing. The main purpose is to obtain the comparison relationship between various weather factors and energy consumption. The analysis of the relationship chart shows that:

(1) The energy consumption relationship is obviously highly negatively correlated with temperature, and highly positively correlated with humidity.

(2) The dew point and UV coefficient have multiple collinearity with temperature so that these two can be discarded. Similarly, cloud cover and visibility also have multiple collinearity with humidity and will be discarded.

(3) The relationship between wind speed and energy consumption is not obvious, but it is not collinear with other weather factors and can be included in the classification process.

According to the above conclusions, we try to construct weather clusters with temperature, humidity, and wind speed. The function function is realized by calling python's own K-means algorithm. The determination of K value will use the elbow rule. As shown in Figure 2, the K value of 3 is more appropriate after analysis.

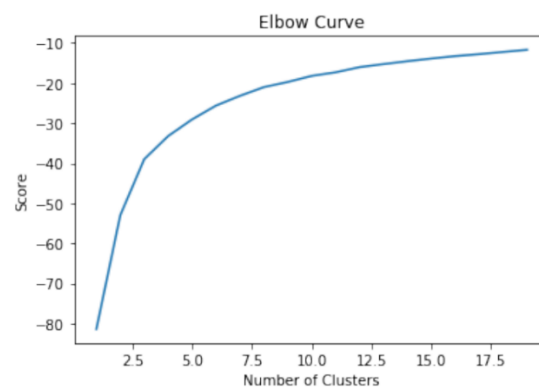


Figure 2: Elbow rule

The K-means algorithm selects the K value of 3, the maximum number of iterations is 600, and adaptively adjusts the model to obtain the scatter clustering effect diagram shown in Figure 3. The clustering result will be added with a new column variable weather cluster to the data set which originally only had two columns of variables which are time and energy consumption.

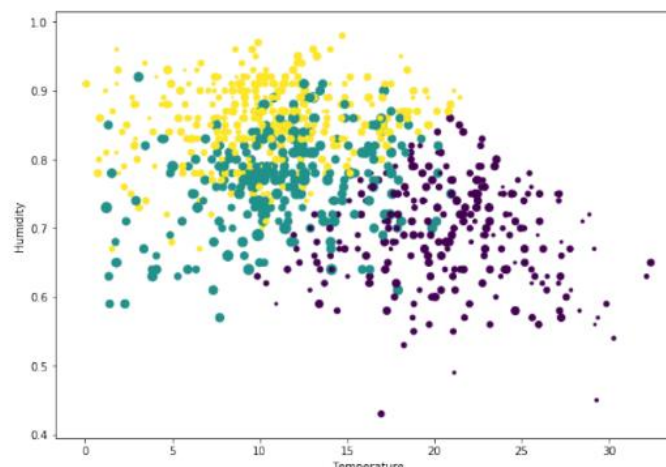


Figure 3: Result of K-means

3.2 Adding holiday information

Generally speaking, when a normal family experiences holidays, their time of staying at home will inevitably be longer than working days, and energy consumption will increase. Therefore, holidays will become special nodes in the forecasting process. The prediction method will obtain unsatisfactory

prediction results at these nodes in lack of the consideration of holiday information which will affect the overall prediction accuracy of the model. Therefore, holiday information will be added to the energy consumption data set as a new column variable. This column of data is 0-1 object data which 1 represents holiday and 0 represents not.

3.3 Building the SARIMAX model

Now except the time information, the input time series data also includes energy consumption, weather clusters, and holiday information. The subsequent process will introduce seasonal parameters to improve the model into SARIMAX model.

According to the processing of time series, stationary detection and smoothing process under non-stationary conditions are required. ADF test is used here. The test results are mainly based on the P value. According to engineering experience, when the P value is less than 0.05, the sequence can be considered stable. According to the result of table 2, the P value is 0.34, which is greater than 0.05. This sequence is considered to be a non-stationary sequence and needs to be smoothed.

Table 2: Stationarity test

Test Statistic	p-value	Number of Observations Used
-1.872794	0.344966	776.00000

Table 3: Stationarity test after smoothing process

Test Statistic	p-value	Number of Observations Used
-6.715004e+00	3.600554e-09	776.00000

The seasonal trend is also an important factor which affects energy consumption. As shown in Figure 4, the energy consumption behavior habits of users in this data set have a relatively obvious seasonal trend, and the energy consumption fluctuates significantly with quarterly changes, so it is necessary to consider the influence of seasonal trends on the prediction effect of the model. The common quarterly difference method to eliminate the influence of seasonal trends is similar to the difference method in the smoothing process. The value obtained by subtracting the previous quarter data from the current quarter data is used as input. This process is to specify the seasonal differential parameters in the SARIMAX mode.

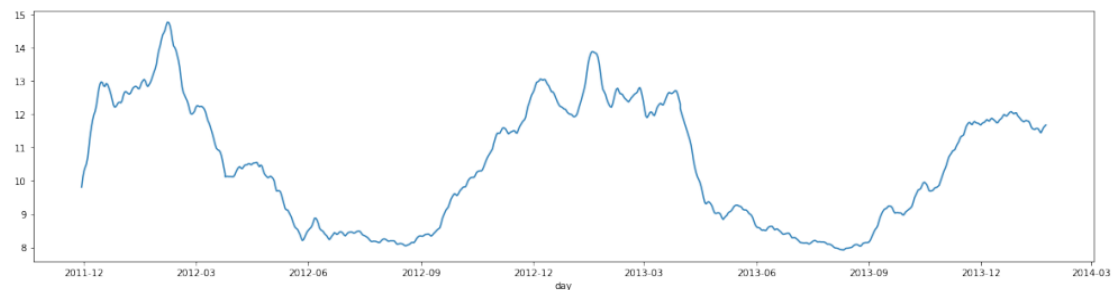


Figure 4: Seasonal trend

After determining all the input variables and realizing the stationary of the time series, this article will determine the core parameters of SARIMAX and formally establish the SARIMAX time series forecasting model. The traditional parameter selection method is the graphical method, that is, draw the autocorrelation (PCF) and partial autocorrelation (PACF) diagrams after the time series difference, and select the corresponding parameters according to the number of peaks in the PCF and PACF diagrams.

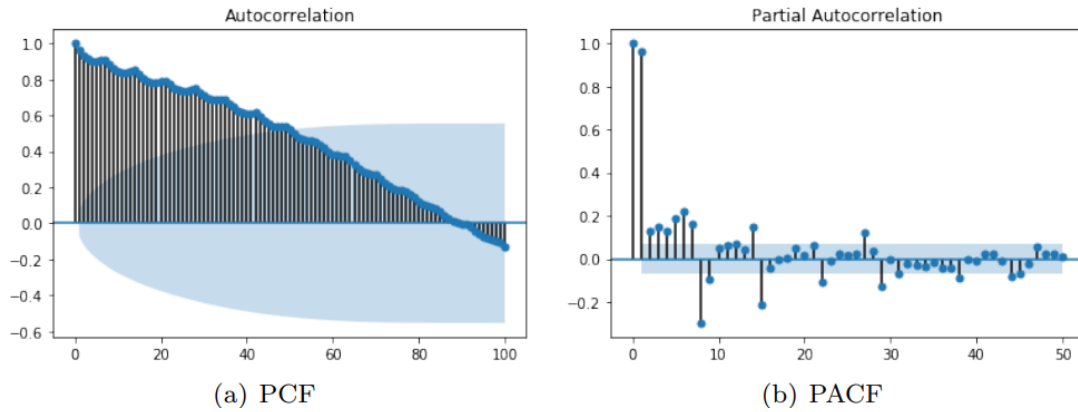


Figure 5: PCF and PACF

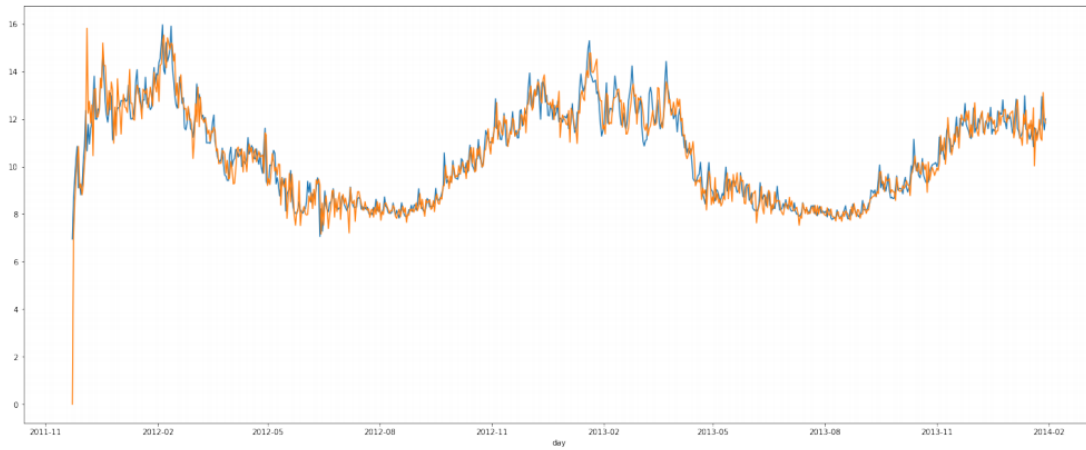


Figure 6: Modeling

The model parameters finally selected by this method are $pdq=(7,1,1)$, $PQD=(1,1,0)$, and the model training and fitting process is shown in Figure 6. The prediction error $MAE=0.5853190227226445$, $MAPE=5.237822685938685$, the prediction effect is shown in Figure 7 (a).

Table 4: Results of grid search

Combination of parameters	AIC
ARIMA(7,0,1)x(1,1,1,12)	1036.919379541391
ARIMA(7,1,0)x(0,0,0,12)	994.8363954367327
ARIMA(7,1,0)x(0,0,1,12)	970.2012844408393
ARIMA(7,1,0)x(0,1,0,12)	1479.304535936925
ARIMA(7,1,0)x(0,1,1,12)	1026.2644211200486
ARIMA(7,1,0)x(1,0,0,12)	954.673303372478
ARIMA(7,1,0)x(1,0,1,12)	953.330595570419
ARIMA(7,1,0)x(1,1,0,12)	1270.9533212153665
ARIMA(7,1,0)x(1,1,1,12)	992.7485287883896
ARIMA(7,1,1)x(0,0,0,12)	992.9622369901334
ARIMA(7,1,1)x(0,0,1,12)	967.8744043767194
ARIMA(7,1,1)x(0,1,0,12)	1472.9288173061918
ARIMA(7,1,1)x(0,1,1,12)	1023.8611553163347
ARIMA(7,1,1)x(1,0,0,12)	953.5732112470148
ARIMA(7,1,1)x(1,0,1,12)	952.3071467226847
ARIMA(7,1,1)x(1,1,0,12)	1268.779109760187
ARIMA(7,1,1)x(1,1,1,12)	992.5456382166236



Figure 7: Comparison of the relationship between weather factors and energy consumption

As we can see, the graphical method is highly subjective, and needs to analyze the results of the parameter selection by itself, which takes more time and does not necessarily have high accuracy. Therefore, this article makes improvements in this area and chooses the grid search method to select the parameters. Grid search refits a new seasonal ARIMA model for each possible parameter combination. Generally, the final parameter combination is obtained by comparing the minimum value in the AIC value. The final forecast effect still needs to be considered comprehensively. The result of grid search is shown in table 4.

3.4 Comparison of the relationship between weather factors and energy consumption

According to the search results, the parameter combination of ARIMA (7, 1, 1) x (1, 0, 1, 12) is selected, and the prediction error is MAE = 0.7994488102003401 and MAPE = 7.272281953582232, and the prediction effect is shown in Figure 7 (b). The analysis shows that the smallest parameter combination of AIC significantly improves the model error, and the prediction effect is not ideal. It can be considered that the parameter combination produces excessive difference.

On this basis, this article has successively tested several adjacent groups of parameter combinations, and finally found that for this group of data sets AIC on condition of taking 1270 up and down can produce a better prediction model. Compare the ARIMA (7, 1, 1) x (1, 1, 0, 12) parameter combination obtained by the graphical method, we find that the ARIMA (7, 1, 0) x (1, 1, 0, 12) parameter combination obtained by the grid search method has better model accuracy, and its prediction error is MAE= 0.577321900617353, MAPE= 5.162511415494644, and the prediction effect is shown in Figure 7 (c).

Analyzing the above experimental results, it is found that only using the SARIMAX model still has a large prediction error, which shows that the predictive ability of a single model is limited. Therefore, this paper uses the output value of the SARIMAX model as the input value of the LSTM model to construct a SARIMAX-LSTM hybrid model, which further improves the prediction accuracy of the model. First, we convert the time series data set into a supervised learning data set, then regularize the data and unify the model dimensions. Then we input the data to train the LSTM model, and finally use the model to make predictions and analyze the prediction results. From the LSTM prediction results in Figure 7 (d), it can be seen that fusing the LSTM model on the basis of the SARIMAX model to build a hybrid model of the two can significantly improve the fitting effect of the model and improve the prediction accuracy.

4. Summary

Aiming at the accurate prediction of regional energy consumption, this paper proposes an energy consumption prediction method based on a SARIMAX-LSTM hybrid model which comprehensively considers weather factors, seasonal factors, and holiday factors. The hybrid model has a strong ability to deal with data fluctuations and a good performance on energy consumption prediction.

At the same time, the model proposed in this paper still has some shortcomings, and the author has also made a corresponding exploration of the direction of improvement: (1) In the process of establishing weather clusters, this paper discards a larger part of it based on multicollinearity, which may lead to incomplete import of weather information. Here, other cluster analysis algorithms with high retention of original data should be used as improvements to improve the utilization efficiency of weather data; (2) Parameters in the SARIMAX modeling process was failed to pass the grid search selection criteria. The over-differential problem that occurs when selecting the smallest combination of AIC parameters needs to be mitigated or eliminated by new methods, such as multiple cleaning of data and adaptation of model structure, etc.; (3) The iterative process of LSTM training converges very fast. The possible reason is that the randomness of the data used for LSTM modeling is significantly reduced after the fitting process of SARIMAX modeling, which may lead to the over-fitting phenomenon of the final model. The possible solutions is to artificially add random data and scale the input data. Future work will focus on the above three deficiencies to retain more valuable information and further improve the prediction accuracy in order to improve the generalization processing capabilities of the model itself for other data sets of the same type.

References

- [1] Fu Zhixin, Li Xiaoyi, Yuan Yue. *Discussion on Key Technologies of Ubiquitous Power Internet of Things [J]. Electric Power Construction*, 2019, 40(5): 1-12.
- [2] Tang Hu, Chen Ailun, Cui Hao et al. *Overview of Urban Energy Internet Research [J]. Hydropower and Pumped Storage*, 2020, 6(1): 13-16.
- [3] Ju Ping, Zhou Xiaoxin, Chen Weijiang et al. *Overview of "Smart Grid+" Research [J]. Electric Power Automation Equipment*, 2018, 38(5): 2-11.
- [4] Gong Jianya, Zhang Xiang, Xiang Longgang et al. *Progress and application of integrated perception and intelligent decision-making in smart cities [J]. Journal of Surveying and Mapping*, 2019, 48(12): 1482-1497.
- [5] Fan Lijun. *Building energy consumption prediction and building energy efficiency analysis based on multiple linear regression model [J]. Natural Science Journal of Xiangtan University*, 2016, 38(1): 123-126.
- [6] Lu Chengju, Li Meng, Liao Baoliang et al. *Energy consumption prediction of diesel passenger cars based on ARIMA model [J]. Shandong Transportation Science and Technology*, 2019, (2): 120-123.
- [7] Guokun Lai;Wei-Cheng Chang;Yiming Yang;Hanxiao Liu. *Modeling Long- and Short-Term Temporal Patterns with Deep Neural Networks [arXiv] [J]. arXiv*, 2017, (0)
- [8] Ji Xinge, Li Hui, Liu Sijia et al. *Short-term photovoltaic power prediction based on MIE-LSTM [J]. Power System Protection and Control*, 2020, 48(7): 50-57.
- [9] XUE Mengfan, HAN Lei, PENG Dongliang. *A combined algorithm of K-means and MTRL for multi-class classification [J]. Journal of Systems Engineering and Electronics*, 2019, (5): 875-885.
- [10] Zheng Yingxin. *Application of cluster analysis based on the elbow rule in data mining in the optimization design of primary and middle school students' travel routes [J]. Electronics World*, 2017, (9): 146.
- [11] You Pin. *Research on the Distribution of DF (ADF) Test Parameters OLS Estimator [J]. Statistics and Decision*, 2009, (20): 20-22.
- [12] Chen Jingjie, Meng Qi. *Prediction of unsaturated airport energy consumption based on two-step decomposition method and SARIMA [J]. Computer Applications and Software*, 2019, 36(4): 46-50, 78.
- [13] Zhang Yan. *The Application of AIC Criterion of Multiple Multiple Linear Regression Model [J]. Modern Trade Industry*, 2016, 37(11): 227-228.