Research and Application of Traffic Forecasting in Customer Service Center Based on ARIMA Model and LSTM Neural Network Model

Mingjie Zhang¹, Yan Sheng², Nuo Tian¹, Wei Liu¹, Hui Wang¹, Longzhu Zhu¹, Qing Xu¹,*

¹State Grid Corporation Customer Service Center, Nanjing, Jiangsu 210000, China
²State Grid Corporation of China Customer Service Center, Tianjin 300300, China

*Corresponding author e-mail: tjxuqing026@sgcc.com.cn

Abstract. Traffic data is the premise of the number of call center seats. The corresponding agents can be arranged for different traffic volumes to achieve optimal configuration of call center human resources. In this paper, ARIMA model and LSTM neural network model based on time series are used to predict traffic. The traffic of the power call center in Hebei Province is taken as an example to conduct experiments on Python software. The results show that LSTM neural network model has higher prediction accuracy than ARIMA model.

Keywords: Traffic Volume, ARIMA Model, LSTM Neural Network

1. Background and Meaning

As the hub of direct contact between customers and enterprises, Customer Service Center uses communication means and computer technology to get better service for customers, to get more extensive customer contact for enterprises, and to publicize the image of enterprises, provides efficient and direct means of service. The operation management of customer service center mainly includes telephone service forecast and scheduling, emergency management, operation site order management, monitoring management and so on. If the traditional empirical scheduling mode is used, not only the high-quality service level cannot be provided, but also the workload of customer service personnel cannot meet the actual production needs. And seat scheduling management is based on traffic forecast. Traffic data is the premise of seat arrangement in call center. Call center can arrange corresponding seats for different traffic. Under the precondition of meeting the service level of the call center, it achieves the optimal configuration of the call center's human resources, and also plays a larger guiding role in the next year's personnel recruitment plan. In order to provide the best service level for the customer service center with the most effective operation cost, it has become a consensus in the customer service center industry to control seat scheduling efficiently and scientifically as a common means of operation management. Reasonable and accurate traffic forecasting not only provides a quantitative basis for scheduling management, but also timely understands business development trends and reduces the loss of corporate image caused by call loss. Therefore, according to the traffic history and related business scenarios, it is particularly important for company operation management...
to predict the changing trend of customer service center traffic and improve the accuracy of traffic prediction.

2. Research Contents
This paper starts with historical traffic data and takes HeBei Province as an example, uses ARIMA model based on time series and LSTM [1] neural network model to predict the traffic. By comparing the average absolute percentage errors of the two models, the optimal model is selected to support the field operation management of customer service center and the optimization of human resources allocation.

3. Traffic Forecast Analysis

3.1 Data Collection and Preprocessing
This paper takes HeBei Province as an example, and selects traffic data from 2018/1/1 to 2019/1/31 as a sample set. The sample set is divided into two parts. The data from 2018/1 to 2018/12/31 are used as training sets to build models, and the data from 2019/1 to 2019/1/31 are used as test samples to build models. The accuracy of the model is verified by comparing with the actual traffic.

In order to improve data quality and eliminate dimension differences, a more accurate model is established. In this paper, traffic data is standardized, and its expression is as follows:

$$x' = \frac{x - \text{average}(x)}{\text{std}(x)} \quad (3-1)$$

Where x is the original data, average (x) is the mean conversion of conversation traffic data, and STD (x) is the standard deviation of traffic data.

3.2 Traffic Forecasting Model Based on ARIMA Model

3.2.1 Introduction to ARIMA Model Principles.
Autoregressive (AR) model, moving average (MA) model, autoregressive moving average (ARMA) model and differential autoregressive moving average (ARIMA) model are famous time series forecasting methods proposed by Box and Jenkins in the early 1970s.

Autoregressive moving average model (ARMA) is a combination of AR and MA, and its expression is as follows:

$$x_t = \phi_1 x_{t-1} + \cdots + \phi_p x_{t-p} + \epsilon_t + \theta_1 \epsilon_{t-1} + \cdots + \theta_q \epsilon_{t-q} \quad (3-2)$$

In the formula above, P represents the autoregressive order; q represents the moving average order; $\phi_1, \phi_2, \ldots, \phi_p$ is the autoregressive coefficient; $\theta_1, \theta_2, \ldots, \theta_q$ is the moving average factor; $x_t$ is a stationary sequence; $\{\epsilon_t\}$ is a white noise sequence with a mean of 0.

Obviously, when p = 0, the ARMA (p, q) model is the MA (q) model. When q=0, the ARMA (p, q) model is an AR (p) model. A lag operator is introduced into the model, which can be simplified as:

$$\Phi(B)x_t = \Theta(B)\epsilon_t \quad (3-3)$$

In the upper formula, $\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$ , $\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$

Autoregressive (AR) model, moving average (MA) model and autoregressive moving average (ARMA) model require that the predicted sequence must be stable. In practice, many series are non-stationary. This introduces the differential autoregressive moving average (ARIMA) model, which is essentially the result of building the ARMA (p, q) model from the stationary series obtained by d-order differential of non-stationary series. The model expression is as follows:

$$\Phi(B)\nabla^d x_t = \Theta(B)\epsilon_t \quad (3-4)$$

Among them,$\nabla^d = (1 - B)^d$
3.2.2 ARIMA Model Modeling Process.

Time series model modeling is usually divided into the following steps:

1. Stationarity test
2. Stationary series are obtained by differential processing of non-stationary time series
3. Determining model order
4. Testing model significance
5. Predicting future values of time series

The time series model modeling flowchart is as follows:

\[
\phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p \\
\theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q
\]

3.2.3 Empirical Study

(1) Stationarity test

When modeling time series data, it is necessary to require that the series be stationary. Otherwise, pseudo regression problems may arise. In this paper, the graph test is used to determine whether the series is stationary. As shown in the figure, it can be seen from Figure 2 that during summer rush hour, the traffic volume is significantly higher than other months, so it can be preliminarily determined that the sequence is non-stationary.
Next, the ADF test is used to test the unit root of the sequence. If the P value is greater than the confidence level alpha, it is proved that there is a unit root in the original sequence, and the sequence needs to be differentiated to ensure the stability of the sequence. The test results are shown in Table 1.

**Table 1. Data stationarity test results**

| variable               | ADF Statistic | 1% critical value | 5% critical value | 10% critical value | result     |
|------------------------|---------------|-------------------|-------------------|-------------------|------------|
| Original sequence      | -2.26         | -3.95             | -3.08             | -2.57             | Nonstationary |
| sequence(-1)           | -5.09         | -4.04             | -3.09             | -2.69             | stationary |

From Table 1, the original sequence is nonstationary, but after first-order difference, it passes the stationarity test. At the same time, it can be seen from Figure 3 that the sequence value after first-order difference fluctuates up and down at 0, so the value of D in the ARMA (p, d, q) model is 1.
(2) Establishing ARIMA model

After the first-order difference, the time series is stationary, d=1 has been determined during the stationary process, and the autoregressive order P and the moving average order q are determined next. Coefficients P and Q are determined by the autocorrelation coefficient (ACF) and partial autocorrelation coefficient (PACF). Different AR (p), MA (q), ARMA (p, q) models can be identified according to different characteristics, as shown in Table 2:

| model          | ACF          | PACF          |
|----------------|--------------|---------------|
| AR(p)          | Trailing     | P-order truncation |
| MA(q)          | q-order truncation | Trailing |
| ARMA(p,q)      | Trailing     | Trailing      |

According to the autocorrelation graph and partial autocorrelation graph, the differences of model structure can be determined as follows:

- ARIMA(3,1,0)
- ARIMA(3,1,1)
- ARIMA(4,1,0)
- ARIMA(4,1,1)

In order to select the optimal model from several initially identified models, AIC criterion is used to determine the order, and the minimum model is the best model. In this paper, using Python software to calculate, it is found that when p=4, q=1, d=1, AIC value is the smallest, so ARIMA (4, 1, 1) model is established for the original sequence.

(3) Model test

The ARIMA (4, 1, 1) model is used to fit time series samples. It verifies the rationality of the model by white noise test on the residuals and provides a basis for determining whether the model is reasonable or not. If the model passes the white noise test, it is proved that the model is reasonable. If the model does not pass the white noise test, it is proved that the model is unreasonable and other models need to be searched.
The residuals of the model are tested for white noise. According to the QQ chart, it can be seen that the residuals sequence of the model satisfies the normal distribution, and the D-W test result value obtained by Python language is 1.98. When the D-W test value is close to 2, there is no autocorrelation, which indicates that the model is good.

![QQ Chart](image)

**Figure 4. QQ**

(4) Model prediction

The ARIMA (4,1,1) model is used to predict the traffic forecasting data of 2019/1/1 to 2019/1/31 in HeBei Province. After restoring the difference data, the forecasted value and the true value polyline are exported as shown in Figure 5.

![ARIMA Model Prediction Chart](image)

**Figure 5. ARIMA model prediction result chart**
The average absolute error percentage of ARIMA (4,1,1) model predictions is 23.32%. It can be seen that the ARIMA (4,1,1) model is not ideal for predicting the traffic in HeBei Province. It is known from the research that the main cause of prediction error is the non-linear characteristics of the traffic data in HeBei Province. ARIMA has a weak ability to describe the non-linear relationship. Although ARIMA model is often used to model and predict the non-linear time series in practical applications, when the non-linear characteristics are strong, the overall prediction effect will be more error. Therefore, in the next chapter of this paper, a non-linear modeling method is introduced to predict the traffic data of HeBei Province.

3.3 Traffic Volume Forecasting Model Based on LSTM Neural Network Algorithm

3.3.1 Introduction of LSTM Neural Network

Neural network is a complex network system, which is connected by a large number of simple basic elements-neurons. It processes information in parallel and converts it nonlinearly by simulating human brain nerve to process information. Recurrent neural networks (RNs) are mainly used for analysis and prediction of time series data. They include input layer, output layer and hidden layer. The value of the hidden layer of RNN is not only related to the input layer at the current time, but also related to the hidden layer at the previous time. However, RNNs retains its historical information transmission time and attenuates. When faced with complex situations, RNNs gradually reduce the ability to remember important features, which to some extent affects the accuracy of prediction. LSTM (Long-Short Term Memory) neural network, as a new and realizable neural network model in recent years, is a special structure of RNNs. It introduces an input gate, an output gate and a forgetting gate, which can solve the problem that RNNN cannot handle long-distance sequences and better coordinate the information distribution in historical storage units. It has time series and selective memory.

(1) Forgotten Gate

LSTM neural networks process sequential data from left to right, here, the forgetting gate determines which information about the cell state is discarded. Expressed mathematically:

\[ f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f) \]  \hspace{1cm} (3-5)

Among them, \( W_f \) and \( b_f \) represent the weight matrix and offset term of the gate, \( x_t \) represents the input vector at time t, and \( h_{t-1} \) represents the output value at time t-1. \( \sigma \) represents the sigmoid function and its range is \([0,1]\). It means that the output of the door is a real vector between 0 and 1, 0 means "completely discarded", and 1 means "fully accepted".

(2) Input gate

The input gate consists of two parts: the first part is sigmoid colon activation function (i.e., the sigmoid function below), the output is \( i_t \), which determines which part needs to be updated; the second part is the \( tanh \) activation function and the output is \( \tilde{C_t} \). The first part of the input gate, output \( i_t \), multiplied by the second part, output \( \tilde{C_t} \), is used to update the cell state. The mathematical expression is:

\[ i_t = \sigma(W_i * [h_{t-1}, x_t] + b_i) \]  \hspace{1cm} (3-6)

\[ \tilde{C_t} = \tanh(W_c * [h_{t-1}, x_t] + b_c) \]  \hspace{1cm} (3-7)

\( W_i \) and \( W_c \) represent the corresponding weights, and \( b_i \) and \( b_c \) represent the corresponding errors.

By expression of the amnesia gate and the input gate, the cell status is updated to \( C_t \), and the mathematical expression is as follows:

\[ C_t = f_t * C_{t-1} + i_t * \tilde{C_t} \]  \hspace{1cm} (3-8)

(3) Output gate
After forgetting the gate and the input gate and updating the unit status, the output gate determines the output information. By processing the latest cell state $C_t$ through $\tanh$ layers and multiplying the result with output vector $o_t$ of the output gate, the hidden state $h_t$ of the final output can be obtained. The mathematical expression is:

$$o_t = \sigma(W_o * [h_{t-1}, x_t] + b_o)$$  \hspace{1cm} (3-9)$$
$$h_t = o_t * \tanh(C_t)$$  \hspace{1cm} (3-10)$$

Where $W_o$ and $b_o$ represents the weight and deviation of the output gate, $h_t$ is the final output value.

### 3.3.2 LSTM Neural Network Modeling Process

![Figure 6. Modeling Flowchart](image)

#### 3.3.3 Empirical Study

1. **Data Refactoring**
   In the LSTM model hidden layer input, the required data matrix format must be three-dimensional, that is, [sample, time step, feature], so we need to reconstruct the standardized data. In this paper, the time step is set to 1, and the feature is set to 1, so that the reconstructed sequence is three-dimensional, matching the input of LSTM neural network.

2. **Model building and adjustment**
   The LSTM model first sets the initialization parameters, then starts the model training until the end of the last training iteration or when the gradient error requirements are met. It is generally believed that it is easier to obtain lower errors by increasing the number of neurons in hidden layers than by increasing the number of hidden layers. The number of hidden layer nodes is directly related to the complexity of solving problems and the number of input and output neurons. [3] If the number of hidden layer nodes is too small, enough connection weight combination cannot be generated to meet the sample learning; if the number of hidden layer nodes is too large, the generalization ability of the network will become worse. In this article, the initial implied layer number is set to one level, the number of implied layer nodes is set to 20, the batch size to 50, the activation function to sigmoid and tanh functions, and the loss function to sequence-loss-by-example function. Learning rate is an important indicator of our participation. Learning rate is one of the most influential superparameters. Compared with other superparameter learning rates, it controls the effective capacity of the model in a more complex way. When the learning rate is the best, the model has the largest effective capacity. [4] The lower the learning rate, the slower the change of loss function. Moreover, the total number of iterations will also affect the prediction accuracy. This paper takes the average absolute percentage error as the model prediction evaluation index.

Average absolute percentage error formula:

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{\hat{\theta} - \theta}{\theta} \right|$$  \hspace{1cm} (3-11)$$
Among them, $\hat{\theta}$ is the traffic forecast value and $\theta$ is the real traffic value.

Through the training of the model, we finally get the learning rate of 0.005, the total number of iterations of 300, and the model accuracy is the highest.

(4) Model prediction
The LSTM model is used to predict the 30-day traffic forecast data in HeBei Province. The results are shown in Figure 7:

![Figure 7. LSTM prediction result graph](image)

The mean square error of the best LSTM neural network model after training is 16.68%, which is less than the average absolute percentage error of ARIMA model. Therefore, from the prediction results, the LSTM neural network in HeBei Province is more accurate than the time series model.

4. Concluding Remarks
Traffic data is the premise of call center seating arrangement. This paper tries to use ARIMA model prediction method based on time series and LSTM neural network algorithm to predict traffic in HeBei province. The advantage of ARIMA model is simple structure, short calculation time, and the disadvantage is that the prediction is not sensitive when data fluctuations are intense. It is suitable for scenarios where traffic data is stable and prediction accuracy is not required. Conversely, the advantages of LSTM model are strong adaptive ability, more accurate prediction of severe data fluctuations, the disadvantage is that the calculation time is longer, and the structure of the model itself is complex. Therefore, compared with time series model, LSTM neural network model has a good prediction effect in Hebei traffic prediction [5-7].

At the same time, there are still some drawbacks in this study. Both models only consider traffic data. In the future, the optimization direction of the models should consider not only increasing policy and business factors, but also modifying traffic at a specific time, but also replacing daily traffic with time-period traffic, and increasing short-term weather forecast to further correct the impact of short-term special weather such as thunderstorms.

Reference
[1] Sepp Hochreiter, Jurgen Schmidhuber.Long short-term memory[J].Neural Computation,1997,9 (8) : 1735- 1780.
[2] Xu Yaoqiang, Fang Leheng, Zhao Donghua, Wang Kai. Large data 2017(18):25-29 for power consumption prediction based on LSTM neural network

[3] Liu C, Wang Y, Kumar K, et al. Investigations on speaker adaptation of LSTM RNN models for speech recognition[C]. IEEE International Conference on Acoustics. 2016.

[4] Li J, Mohamed A, Zweig G, et al. LSTM time and frequency recurrence for automatic speech recognition[C]. Automatic Speech Recognition & Understanding. 2016

[5] Cheng Xiaopei; Analysis of Money Supply Variation Based on ARIMA Model[J]; Cooperative Economics and Technology; 2018-14

[6] Lou Pengwei; Wu Xiufeng; Zhang Xueliang; Xu Jiabo; Wang Kai;; Analysis of Xinjiang Brucellosis Epidemic Trend Based on ARIMA Product Seasonal Model[J]; Journal of Xinjiang Medical University; 2017-01

[7] Cheng Yun; Cheng Xiaogang; Tan Miaomiao; Zhou Kai; Li Haibo;; Traffic flow prediction based on ARIMA and wavelet neural network combined model[J]; Computer Technology and Development; 2017-01