Enterprise Economic Forecasting Method Based on ARIMA-LSTM Model

Xiaofei Dong[1](✉), Xuesen Zong[1], Peng Li[2], and Jinlong Wang[1]

[1] Qingdao University of Science and Technology, Qingdao, China
unsolvedcila@163.com
[2] Qingdao Yilian Information Technology Co. LTD, Qingdao, China

Abstract. Enterprise economic forecast is an important part of the development of enterprises, which can help the government to judge the development of enterprises quickly and effectively so as to make scientific decisions of China. With the development of Internet of Things (IOT) technology, enterprise's IOT data can bring strong data basis to enterprise's economic forecast. In order to obtain more accurate results of enterprise economic forecasting, a method of enterprise economic forecasting based on Auto regressive Integrated Moving Average and Long Short Term Memory networks (ARIMA-LSTM) model is proposed, which solves the problem that a single forecasting algorithm can only predict according to a single economic development data. The model uses ARIMA model to predict the linear data of time series such as IOT data, and LSTM to predict the nonlinear relationship. Combined with the historical economic data of enterprises, ARIMA-LSTM model is used to predict the future economic development of enterprises. Comparing the prediction results with ARIMA model and ARIMA-LSTM model without IOT data, it is found that the model has the smallest RMSE, MAE and MAPE. The results show that the model can effectively predict the economic situation of enterprises.

Keywords: Enterprise economic · IOT · ARIMA · LSTM · Forecast

1 Introduction

1.1 Research Significance

With the advent of the second Centenary Goal, many local governments have issued plans to support the acceleration of the development of leading enterprises and formulated development goals for leading enterprises. The State Council has also issued a number of opinions on promoting industrial development, guiding the development of emerging industries, expanding existing high-tech enterprises and fostering small and medium-sized technology-based enterprises. The government needs to monitor the development of enterprises and timely adjust the list of enterprises to be cultivated. To predict and estimate the specific economic status and future development trend of enterprises can help government personnel to have a clearer understanding of the economic development trend of enterprises, provide detailed information and data support

Z. Lv and H. Song (Eds.): INTETAIN 2021, LNICST 429, pp. 36–57, 2022.
https://doi.org/10.1007/978-3-030-99188-3_4

for decision-making, and ensure the accuracy of decision-making [1]. The operating income of an enterprise is one of the financial indicators that directly measure the scale of enterprise operation and indirectly reflect its operating results [2]. The management of enterprise operating income is an important aspect of enterprise financial management. However, it is difficult to judge the future development trend of an enterprise only by relying on the data reported by the enterprise, and a series of problems such as excess capacity and insufficient competitiveness of the enterprise may occur if the wrong large-scale investment is made. Therefore, it is of great significance for the government to monitor and manage enterprises and regions to study the future economic development trend of enterprises.

1.2 Research Background

Many scholars have conducted in-depth research on enterprise economic forecasting. Xiao et al. evaluated the performance of enterprise investment decision by Back Propagation (BP) neural network [3]. Wang et al. proposed an enterprise economic risk prediction model based on big data fusion to predict enterprise economic risk [4]. These methods are of great help to scientific decision making of China.

However, the above methods all use historical economic data for linear regression prediction. The factors affecting the economic development of enterprises are diversified, and the business operation and related data management can never be explained by a simple linear regression system. With the continuous development of learning technology, more and more scholars tend to use nonlinear regression model to model and analyze financial time series. Barker et al. constructed a unified modeling framework, adjusted the Machine Learning model hyperparameters and generated the prediction interval by using time series cross-validation, with good results [5]. Hryhorkiv et al. used artificial neural network to predict the nonlinear part of stock fluctuations and achieved good prediction results [6]. Feng et al. studied economic prediction based on wavelet neural network, which is used to predict nonlinear systems [7]. Huang et al. used support vector machine and other three different models to predict enterprise operating income data and evaluated the predictive ability of the models [8]. However, the nonlinear prediction are according to the nonlinear consolidation of economic data, without considering the other factors may be the effects on the economy, according to the analysis and forecasting of economic data, only can't analysis because the economic impact of other objective factors, lead to prediction has limitations, is not conducive to scientific decision-making of the government. In view of the method of using multiple factors to predict the economy, some scholars put forward the prediction model based on the traditional BP neural network and time series to predict the sales of the enterprise in the future unit time [9], but the training data are independent and cannot be combined with historical data for training. Wu used Long Short Term Memory networks (LSTM) model to forecast sales [10], avoiding the independent prediction problem of BP network and the long-term dependence problem of recurrent neural network. However, the model can only be used after obtaining all the data of influencing factors. If the data of current influencing factors has not been generated or missing, it cannot be predicted.

In view of the problems existing in the above existing methods, this paper will conduct in-depth studies on factors that can reflect the economic development status of enterprises. Internet of Things (IOT) data can truly reflect the characteristics of the operation process of enterprises [11–13], and excellent performance of Auto regressive Integrated Moving Average (ARIMA) in IOT data prediction [14–16]. A forecasting model of enterprise economic development based on ARIMA-LSTM is designed and implemented. The IOT data combined with the data of economic impact factors of Internet enterprises are used to train the economic development of enterprises. The ARIMA model is used to predict the changes of the data of influence factors in the future, and LSTM is used to integrate the data of historical economic data non-linear. The experiment shows that the model can accurately predict the future economic development trend of enterprises, and help government personnel to accurately control and timely adjust enterprises.

2 Correlative Knowledge

2.1 The Relationship Between IOT and Economic Development

The concept of the IOT was first put forward by the Massachusetts Institute of Technology in 1999. With the continuous development of information technology, the IOT is more and more widely used. The IOT can not only provide real and ordered data, but also bring great convenience to data calculation and analysis. Structured data of the IOT is also more convenient for researchers to sort out and study data, from which more valuable information can be found. And since IOT data is machine log data generated by networked devices, it is generally not allowed and does not need to be modified. For government departments, it avoids the situation that enterprises report false data, and the data is more real and reliable.

Many scholars have studied the connection between the IOT and economic development. Xu et al. analyzed the relationship between regional economic development and electricity consumption data and predicted the Gross Domestic Product growth rate of Jiangsu province with BP neural network algorithm, which achieved good results [11]. Wang et al. analyzed the connection between the IOT and the development of regional economy and proved that the IOT and regional economic development have a very strong correlation [12]. Chen et al. analyzed the significance and role of the IOT in economic development [13]. Research shows that the IOT data is closely related to economic development, and the economic development of enterprises can also be intuitively reflected through the IOT data.

2.2 Application of Time Series Model in IOT Data Prediction

IOT data is the data generated during the operation of equipment. The data must be time series data with time stamps, which is very important for the calculation and analysis of data and can be processed by using time series analysis algorithm. Moreover, the data flow of the IOT is stable and predictable, which makes the data of the IOT more suitable for stable prediction [17].

Many scholars applied the time series prediction model to the data prediction of the IOT. Tang et al. analyzed the practicability of three commonly used models in the prediction of three different types of parking lots, and the results proved that Autoregressive

Integrated Moving Average model (ARIMA) and Back Propagation (BP) neural network had good accuracy models in short-term prediction [14]. Lin used ARIMA model of geometric empirical mode decomposition to predict electricity sales and proved that ARIMA model has good prediction accuracy for electricity sales data [15]. Wang et al. used two forecasting methods combining BP neural network, ARIMA and Support Vector Regression (SVR) to predict the electricity consumption of 20 energy-intensive enterprises by taking into account macroeconomic indicators, upstream and downstream product output, and weather [16]. The experimental results show that ARIMA+SVR performs better in predicting power consumption. The above experiments show that the time series model can get good prediction effect in predicting the data of IOT, but only predicting the data of IOT is not enough to predict the economic development of enterprises, so it is necessary to redesign the method of predicting the economic development of enterprises.

3 Enterprise Economic Forecasting Method Based on ARIMA-LSTM Model

The overall process of the enterprise economic forecasting method based on ARIMA-LSTM model proposed in this paper is shown in Fig. 1. It is mainly composed of several ARIMA models and one LSTM model. The ARIMA model is mainly used to predict the original influencing factor data of each time series and get its predicted value respectively. Then, the training data, predicted value and other influencing factors are combined and normalized. The LSTM model is used to construct the functional relationship between the combined influencing factors and enterprise operating income, and the final enterprise operating income data is predicted.

3.1 Enterprise Development Environment

There are many factors related to economic development in the process of enterprise development. Faced with numerous data, data can be divided into Internet data and IOT data according to data sources. In the Internet data, the economic development of enterprises is closely related to the information of operating income of enterprises, the news and public opinion of enterprises, the number of highly educated talents. In the IOT data, the number of vehicles in the parking lot data, enterprise electricity consumption data, enterprise access control personnel information data, working hours data reflect the personnel situation of enterprises. Energy consumption and personnel activity can also reflect the economic development of enterprises.

When combining the IOT data to predict the enterprise economy, time series analysis and neural network model are used alone to predict accompanying by the following shortcomings. Time series analysis cannot predict the influence of other factors on the results. Neural network model can predict the results according to the existing influencing factor data, but can not predict the future results when there is no influencing factor data. In this paper, ARIMA which has good prediction results in time series analysis and LSTM which is suitable for long term series prediction in neural network are selected. Combined with the characteristics of the two, the time series influencing factor data and enterprise economic data are predicted, which can realize the enterprise economic forecast combined with the IOT data.

3.2 Model Principle

As mentioned above, the IOT data can accurately reflect the operating status of enterprise equipment, which can be predicted by the time series model, and the combined model can solve the limitation of single model prediction, so as to effectively predict enterprise economy.

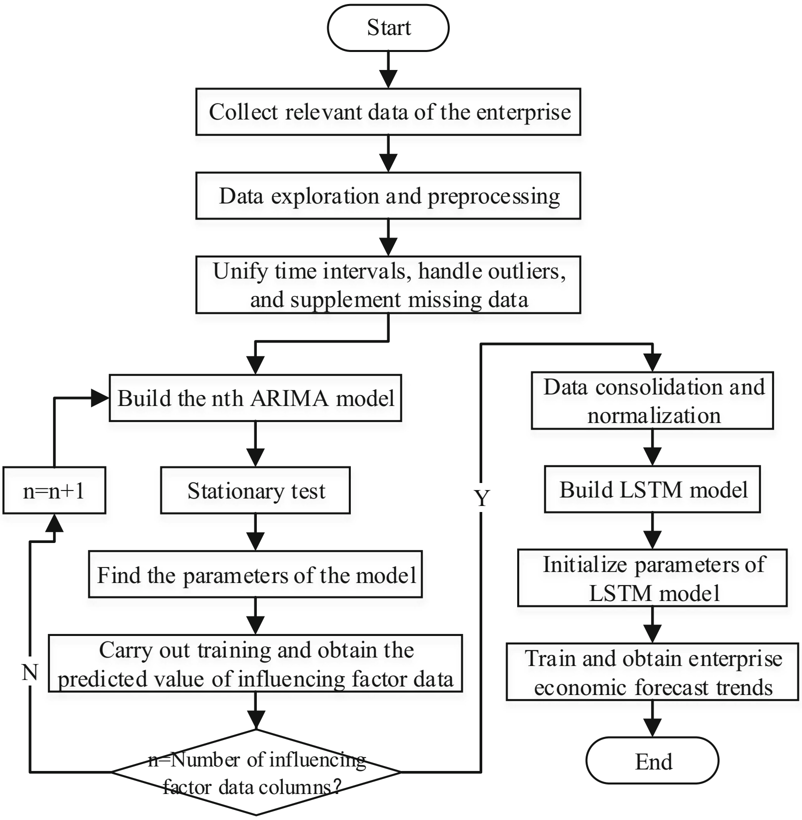


Fig. 1. The overall process of enterprise economic forecasting method based on ARIMA-LSTM model.

Table 1 shows the parameters of the main related variables of the model.

In most studies, the model is fixed and the validity is evaluated for different input characteristics, but there is a lack of research on hyperparameters [4]. After study, the model sets the data set of influencing factors as $F_t = \{O_{MR}, G_{MR}, H_t, P_t, E_t, N_t, W_t, m, P_{MR}, E_{MR}, N_{MR}\}$. Among them, O_{MR}, G_{MR}, P_{MR}, E_{MR}, N_{MR} is the lag data with the greatest correlation. In the process of forecasting the development of enterprise economy, the influence of data on enterprise economy also has certain characteristics. Generally, enterprise economy will not immediately show the change of enterprise economy with the change of data in the month, but there is a certain lag. Therefore, the most relevant IOT data are added into the influencing factors. Due to the absence of time series in Internet data, it is impossible to predict future data based on existing data. Therefore, the historical Internet data with the greatest correlation is added to the influential factor data set to predict enterprise operating income data. In addition, the data related to the current time is added to ensure the latest data. The current month data can be used to find the time characteristics of time series data.

Table 1. Variable parameter table.

Parameter	Parameter meaning
I^E	General term for Internet data
I^{OT}	General term for IOT data
M_t	The operating income data of the enterprise
O_t/O_{MR}	T time/maximum relevance of enterprise news public opinion information data
G_t/G_{MR}	T time/maximum relevance of enterprise positive information data
H_t/H	T time/all time data of highly educated talents
$P_t/P/P_{MR}$	T time/all time/maximum relevance of the number of vehicles in the parking lot
$E_t/E/E_{MR}$	T time/all time/maximum relevance of enterprise electricity consumption data
$N_t/N/N_{MR}$	T time/all time/maximum relevance of enterprise access control personnel information data
W_t/W	T time/all time working hours data
m	Month data
F_t/F	T time/all time general name of influencing factor data
F_{pre_n}	Influencing factor data for the nth month in the future

ARIMA model is responsible for predicting each time series data. Based on historical data, it predicts the vehicle access data of parking lot in the next year, electric power consumption data of enterprises, information data of access control personnel of enterprises, number of highly educated talents, working hours, etc.

Assume that the time series running data of an enterprise includes H, P, E, N and W.

The calculation formula of ARIMA model is as follows.

$$y_t = \sum_{m=1}^{p} \varphi_m y_{t-m} - \sum_{j=1}^{q} \theta_j a_{t-j} + a_t + c \tag{1}$$

Where $\varphi_m(m = 1, 2, ..., p)$ is the coefficient of the autoregressive model, $\theta_j(j = 1, 2, ..., p)$ is the average sliding coefficient, a_t is the white noise sequence, the mean value is 0, and c is a constant. The enterprise time series operation data H, P, E, N, W is used to obtain the prediction data results of time series through ARIMA model. Take parking lot data P as an example, if the white noise of parking lot time series is ε_t.

$$P_t = \sum_{m=1}^{p} \varphi_m P_{t-m} - \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t + c \tag{2}$$

The obtained prediction sequence is recorded as $P_{t+n}(n = 1, 2, ...)$.

The prediction results of each column of enterprise time series operation data in F in the next n months predicted by ARIMA model are recorded as H_{t+n}, P_{t+n}, E_{t+n}, N_{t+n} and W_{t+n} respectively. Combining the monthly data and the maximum lag data, the predicted influencing factor data set F_{pre_n} is obtained.

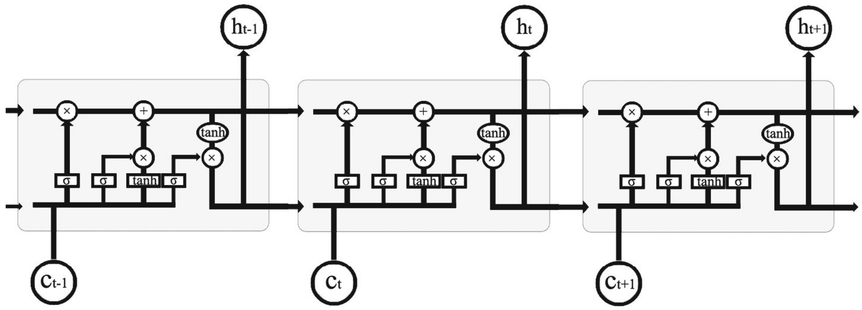


Fig. 2. LSTM model structure.

$$F_{pre\_n} = \{O_{MR+n}, G_{MR+n}, H_{t+n}, P_{t+n}, E_{t+n}, N_{t+n}, W_{t+n}, m_{t+n}, P_{MR+n}, E_{MR+n}, N_{MR+n}\} \tag{3}$$

The structure of LSTM model is shown in Fig. 2. Normalize the influencing factor data set F_t, prediction influencing factor data set F_{pre_n} and enterprise operating income data M_t as the input of LSTM model to make a nonlinear prediction of the enterprise's future economy. The prediction process is as follows:

Step 1: The forgetting gate reads the output of the last cell h_{t-1} and the input of the current cell c_t, and the sigmoid function σ outputs a value between 0 and 1. 1 means "keep completely" and 0 means "discard completely".

$$forget_t = \sigma(W_f \cdot [h_{t-1}, c_t] + b_f) \tag{4}$$

Step2: A sigmoid layer determines what information needs to be updated, and a tanh layer generates content $\tilde{C}_t$ for updating.

$$i_t = \sigma(W_i \cdot [h_{t-1}, c_t] + b_i) \tag{5}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, c_t] + b_C) \tag{6}$$

Then, update C_{t-1} to C_t. Multiply the old state by $forget_t$, discard the information to be discarded and add $i_t \cdot \tilde{C}_t$ as the new candidate.

Step3: Run a sigmoid layer to determine which part of the cell state will be output. The cell state is processed through tanh (to get a value between −1 and 1) and multiplied by the output of the sigmoid layer, and output the part that determines the output.

$$out_t = \sigma(W_o[h_{t-1}, c_t] + b_o) \tag{7}$$

$$h_t = out_t * \tanh(C_t) \tag{8}$$

It can be seen from the formula that the prediction results of LSTM model are closely related to historical data and stored by storage units. LSTM avoids the problem of gradient extinction and gradient explosion, and can still learn long-term dependence from corpus. Therefore, more accurate results can be obtained by using LSTM model to predict enterprise economic development sequence.

The prediction process of ARIMA-LSTM combined model is as follows:

$$forget_t = \sigma(W_f \cdot [h_{t-1}, \{F_t, F_{pre\_n}, M_t\}] + b_f) \tag{9}$$

$$i_t = \sigma(W_i \cdot [h_{t-1}, \{F_t, F_{pre\_n}, M_t\}] + b_i) \tag{10}$$

$$\tilde{C}_t = \tanh(W_C \cdot [h_{t-1}, \{F_t, F_{pre\_n}, M_t\}] + b_C) \tag{11}$$

$$out_t = \sigma(W_o[h_{t-1}, \{F_t, F_{pre\_n}, M_t\}] + b_o) \tag{12}$$

$$h_t = out_t * \tanh(C_t) \tag{13}$$

The optimization objective of ARIMA-LSTM combined model is to minimize the sum of the gap between the 16-month actual value E and the predicted value A in the predicted results:

$$Min \sum_{i=1}^{12} \left| \frac{A_i - E_i}{E_i} \right| \tag{14}$$

3.3 Model Process

The modeling process is mainly divided into two steps:

1) Data preparation

The model data preparation process is shown in Fig. 3.

Firstly, collect relevant data of the enterprise from relevant departments, sort out I^E, I^{OT} and M_t that may be associated, search for O_{MR}, G_{MR}, P_{MR}, E_{MR} and N_{MR}, preprocess the data, use 3sigma principle to find and delete outliers, and use multiple interpolation method to supplement missing data.

Considering the lag of enterprise economic data, we calculated the prediction effect of O_t, G_t, P_t, E_t and N_t data with one month lag, two months lag and three months lag respectively, and selected the forecast data with the best prediction effect as O_{MR}, G_{MR}, P_{MR}, E_{MR}, N_{MR}.

The results of correlation analysis by using Corel function of Excel are shown in Table 2.

As can be seen from the above table, the correlation of data with a lag of one month is the highest, so it can be concluded that $O_{MR} = O_{t-1}$, $G_{MR} = G_{t-1}$, $P_{MR} = P_{t-1}$, $E_{MR} = E_{t-1}$ and $N_{MR} = N_{t-1}$. Data set F_t of influencing factors can be expressed as $F_t = \{O_{t-1}, G_{t-1}, H_t, P_t, E_t, N_t, W_t, m, P_{t-1}, E_{t-1}, N_{t-1}\}$.

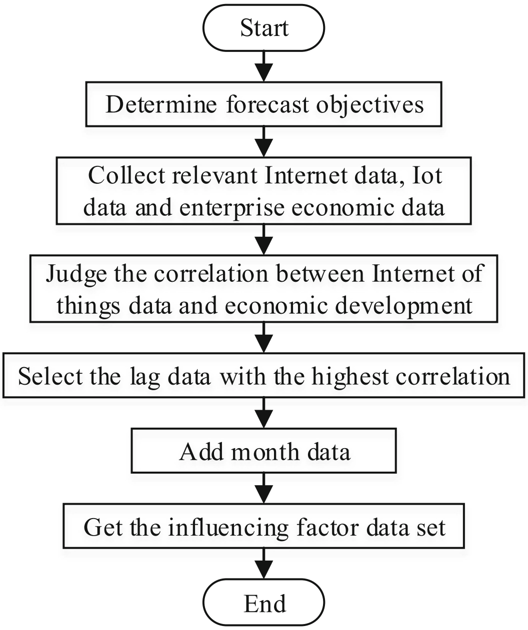


Fig. 3. Model data preparation process.

2) Construct ARIMA-LSTM combination model

The specific process of constructing the combined ARIMA-LSTM model is shown in Fig. 4.

Step1: Select a kind of time series data and test the stationarity of the time series. If the sequence is unstable, the difference operation is carried out to improve the stationarity until the time series meets the stationarity requirements. The number of differences is called d.

Table 2. Correlation between lag time and five kinds of data.

Correlation	O_t	G_t	P_t	E_t	N_t
Raw data	0.147	0.139	0.618	0.383	0.192
One month behind	0.183	0.169	0.911	0.712	0.192
Two month behind	0.122	0.131	0.424	0.033	0.168
Three month behind	0.101	0.128	0.151	−0.244	0.138

Step2: Find the parameter (p, d, q) that is most suitable for ARIMA model. Firstly, according to the number of differences, the number of differences of the model can be determined as d, thus transforming the problem into determining parameter C of the ARMA model. Then, we use the Akaike Information Criterion (AIC) minimum information criterion [18] to limit the range of p and q, and substitute the traversal mode of (p, q) into the AIC criterion to find a combination that minimizes AIC value. This is the optimal combination of parameters for the model.

Step3: The ARIMA model was constructed and trained according to the optimal combination parameter (p, d, q), and the prediction results in the next 16 months were obtained.

Step4: For time series data H_t, P_t, E_t, N_t, and W_t, construct ARIMA model respectively and obtain the predicted value H_{t+n}, P_{t+n}, E_{t+n}, N_{t+n} and W_{t+n} of each data. Combine monthly data and maximum lag data to get the data set of forecast influencing factors:

$$F_{pre\_n} = \{O_{t-1+n}, G_{t-1+n}, H_{t+n}, P_{t+n}, E_{t+n}, N_{t+n}, W_{t+n}, m_{t+n}, P_{t-1+n}, E_{t-1+n}, N_{t-1+n}\} \tag{15}$$

Step5: Merge influencing factor data F_t, forecast result data F_{pre_n} and enterprise business revenue data M_t, and represent them in the form of array: overall data set $X = [x_0, x_1, ..., x_{95}]^T$, training set $X_{train} = [x_0, x_1, ..., x_{83}]^T$, and test set $X_{test} = [x_{84}, x_{85}, ..., x_{95}]^T$. In order to unify the weight of data, G is normalized, which can retain data characteristics to the maximum extent and better fit the non-linear relationship between various data and enterprise operating income. In this paper, the maximum and minimum normalization method is adopted to normalize the data:

$$x_{norm} = \frac{x - x_{\min}}{x_{\max} - x_{\min}} \tag{16}$$

Where, x represents sample data; $x_{\max}$ and $x_{\min}$ represent the maximum and minimum values in training data or test data respectively. Data of training set and test set after normalization are denoted as X_{train1} and X_{test1} respectively.

Step6: Initialize LSTM model parameters and build training network. The normalized training set X_{train1} is input into the network for training, and then the output result in the next month is obtained as the predicted value of enterprise economy.

Step7: Iterate the actual value of enterprise development on the model data and re-conduct the model process, and obtain the predicted value of the next 16 months successively as the final prediction result.

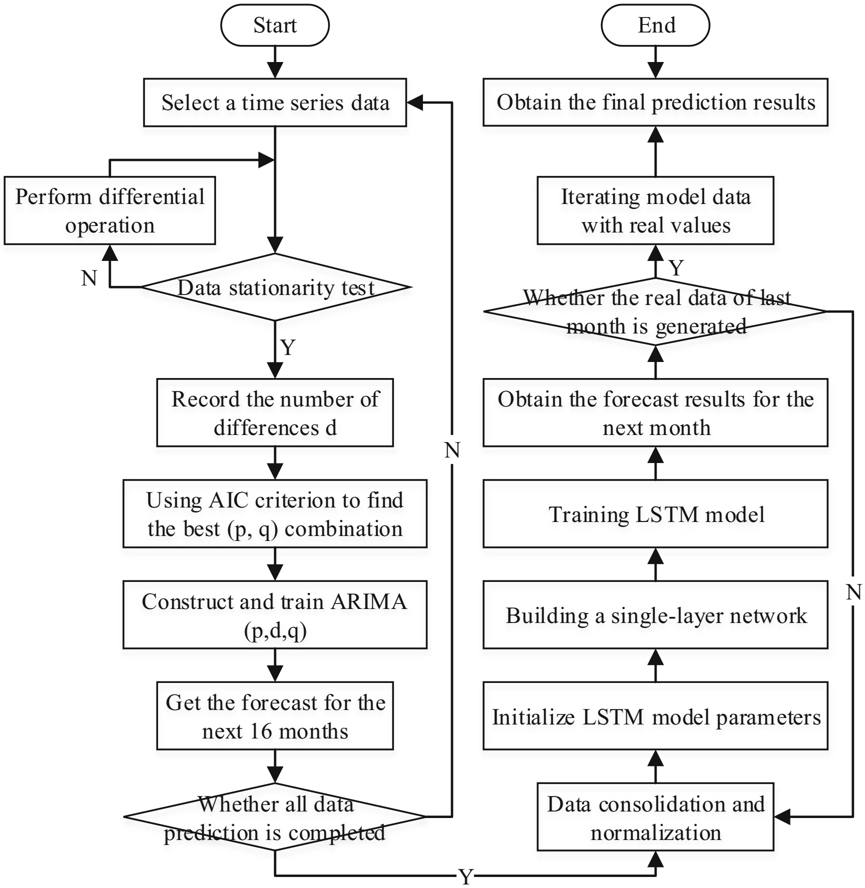


Fig. 4. The process for building a composite ARIMA-LSTM model.

ARIMA-LSTM combined model solves the limitation that ARIMA model can't make nonlinear prediction for a variety of influencing factors and LSTM model can't predict the situation without influencing factor data. Combined with the reflection of IOT data on enterprise economy, the model can predict the change of economic trend caused by emergencies. It has realized the more accurate forecast to the future enterprise economic trend.

Algorithm 1 gives the specific ARIMA-LSTM combination algorithm.

Algorithm 1 ARIMA-LSTM combination algorithm

Input: IOT data set, enterprise historical economic data, relevant Internet data, time data.
Output: Business economic forecasts for the next 16 months.

```
1：//Step 1: Data stationarity test.
2：data = ADF(train) // Data stationarity test
3：d = 0
4：if data >= 0.05
5：  d = d + 1;
6：  data = ADF(train).diff(d);
7：end if
8：//Step2: Traverses to find the best parameter.
9：for each
10：  for each
11：    temp = ARIMA(train, (p, d, q))
12：    AIC = AIC(temp)
13：  end for
14：end for
15：p, q = idxmin(AIC)
16：//Step3: Construct ARIMA model and predict each column time series value.
17：model = ARIMA(train, (p, d, q))
18：forecast1 = model. forecast(16)
19：//Step4: Repeat step1-Step 3 to predict other time series data.
20：forecast = [forecast1, forecast2, …, forecast16]
21：//Step5: The predicted results are combined and normalized with the original training set.
22：trainData = train, forcast
23：norm = Normalization(trainData) // Normalize the data
24：//Step6: Create the LSTM model, initialize the parameters, and get the predicted values.
25：create LSTMmodel;
26：select the best parameters of LSTM;
27：predict = LSTM(trainData); // Get the predicted value for the next month.
28：//Step7: Iterate the data with true values and repeat Step22-Step27 for prediction.
```

4 Experiment and Analysis

4.1 Experimental Environment

The hardware environment of this experiment is a laptop computer equipped with I5-8250U CPU and 8 G memory. The software environment is Windows 10 operating system, the programming language is Python 3.7.1, and the LSTM neural network is constructed using Tensorflow deep learning framework.

4.2 Data Set Selection

The data set used in this experiment is from the data of an office building enterprise in Qingdao. Three IOT data, namely the number of parking lots, electricity consumption and working hours, and four Internet data including the total number of employees, the number of highly educated talents, the number of favorable enterprises and the number of online public opinions, are selected as well as the economic data of enterprises. A total of 96 groups of data were selected from January 2013 to December 2020.

Table 3. Some examples of training data of enterprise economic forecasting.

Time	O_{MR}	G_{MR}	H_t	P_t	E_t	N_t	W_t	m	M_t
2013/1/1	0	0	0	5575	5720	892	248	1	717.75
2013/2/1	0	0	0	3189	1524	892	224	2	628.24
2013/3/1	0	0	0	5393	4330	892	248	3	565.05
…	…	…	…	…	…	…	…	…	…
2016/11/1	0	0	9	5687	4239	1099	240	11	710.13
2016/12/1	0	0	9	5617	5426	1099	248	12	675.87
2017/1/1	0	0	11	5553	5623	1122	248	1	717.42
2017/2/1	0	0	11	3297	1557	1122	224	2	639.88
2017/3/1	0	0	11	5424	4337	1122	248	3	550.47
2017/4/1	0	0	10	6780	4287	1102	240	4	696.55
…	…	…	…	…	…	…	…	…	…
2019/10/1	0	0	27	7245	3981	1287	248	10	761.59
2019/11/1	0	0	27	5756	4136	1287	240	11	723.40
2019/12/1	0	0	27	6135	5558	1287	248	12	691.58

First of all, the data set is preprocessed to remove outliers in each column according to the most commonly used 3sigma principle [19] in anomaly detection, and the missing values are supplemented by multiple interpolation methods. After pre-processing, data from December 2019 and before were used as training sets, and data from the whole year of 2020 and the first half of 2021 were used as test sets. Since enterprise economic data is related to time series and other factors, and IOT data has a lag effect on the development of enterprise economy, enterprise economy will not immediately show a relationship with IOT data. Therefore, in order to obtain more accurate prediction results, we took the IOT data of the last month with the highest correlation as the input parameter of the model, and added them to the training of LSTM model after pretreatment, which can improve the prediction accuracy of LSTM. The processed data of the combined model are shown in Table 3.

The line chart of data of each influencing factor is shown in Fig. 5. It can be seen from the figure that parking data and electricity consumption data change with time with obvious regularity and have certain time series characteristics.

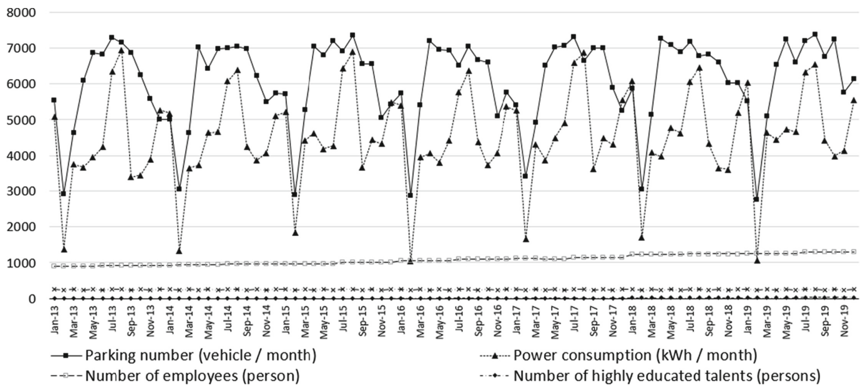


Fig. 5. The process for building a composite ARIMA-LSTM model.

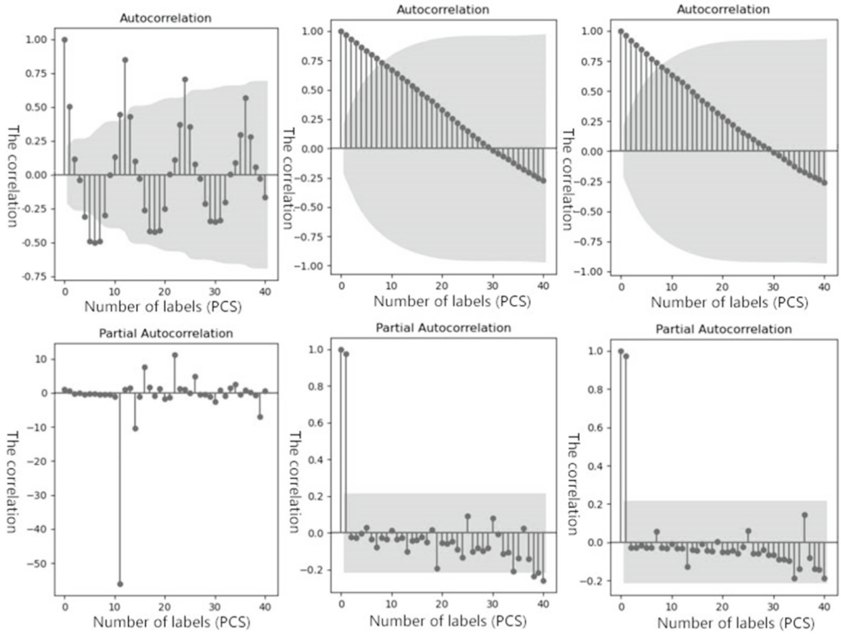
4.3 Experimental Process

ARIMA-LSTM combined model predicts the influencing factor data of each column of time series to obtain the time series forecast data in the next 16 months. Taking parking data as an example, $P_{train} = \{P_1, P_2, ..., P_{84}\}$ is used as training data and $P_{test} = \{P_{85}, P_{86}, ..., P_{102}\}$ is used as test set. ARIMA model is used to train the training data. In order to eliminate the adverse effects of pseudo-regression of time series model, stationarity test of data is required. The most commonly used methods for stationarity testing are Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analysis. ACF and PACF diagrams of vehicle access data (a), enterprise electricity consumption data (b), and enterprise access control personnel information data (c) are shown in Fig. 6. Due to the limitation of space, we only take the total number of vehicles visited by enterprises in one month as an example to introduce the process and results of stationarity test in detail.

As can be seen from Fig. 6 (up (a)), the ACF graph fluctuates up and down instead of rapidly approaching 0, which is trailing. In addition, the Augmented Dickey-Fuller (ADF) test result $P = 0.38 > 0.05$ accepts H0, which also indicates that the sequence is non-stationary and difference is required to make the sequence stable. The data sequence diagram after first-order difference is shown in Fig. 7. It can be seen that the sequence fluctuates stably around 0. At this point, ADF test result $P = 0.004 < 0.05$ rejects H0, indicating that the sequence is stable and can be used for prediction of ARIMA model.

After ARIMA's prediction, the prediction result of parking data in the nth months in the future is denoted as P_{t+n}. Similarly, other sequences were input into ARIMA model to obtain the prediction results of all influencing factor data in the nth months in the future, which were denoted as H_{t+n}, P_{t+n}, E_{t+n}, N_{t+n} and W_{t+n} respectively. Data set F_{pre_n} of influencing factors was obtained by combining monthly data and maximum lag data.

After the data is merged and normalized, the LSTM model is created and parameters are set. Initialize the LSTM argument using Python's Keras library. The activation function of LSTM module was tanh, the fully connected artificial neural network receiving LSTM output was set as Linear, and the rejection rate of each network node was set



(a) Vehicle access data (b) Electricity consumption data (c) Access Control Personnel data

Fig. 6. ACF and PACF graphs of IOT data.

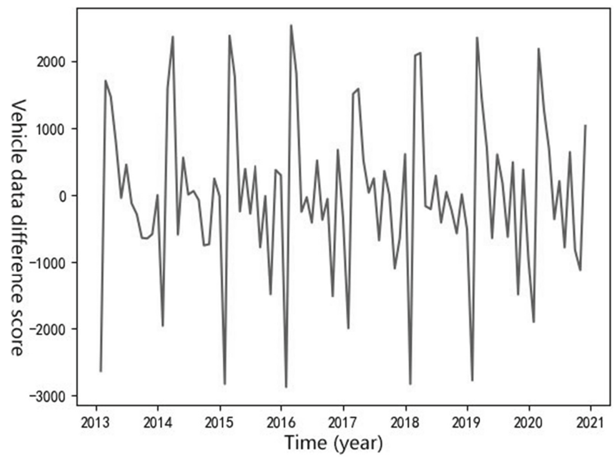


Fig. 7. Vehicle data diagram after first order difference.

as 0.2. RMSprop algorithm suitable for RNN is used for iterative updating of mean square error and weight parameters. The epoch interval for model training was set to 100 and the Batch size was set to 500. Train and get a forecast for the next month. As the test data of one and a half years is set, the results of the next month are continued to be predicted after the iteration results of the true value of the current month, until the enterprise economic data of the next year and a half is predicted as the final model prediction result.

4.4 Results and Analysis

In order to reflect the advantages of the combination model combined with IOT data in enterprise economic forecasting, we choose a single ARIMA model, a single LSTM model and a combination model without IOT data to compare the enterprise forecasting results.

In this paper, Root Mean Square Error (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE) are selected as the calculation of prediction accuracy.

1) RMSE

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_t - y_t)^2} \tag{17}$$

Where $\hat{y}_t$ represents the predicted value of time t, y_t represents the actual value of time T, and n represents the total predicted time. The RMSE range is $[0, +\infty)$, and the smaller the value is, the higher the prediction accuracy is. When the predicted value is in perfect agreement with the real value, the value is 0. The greater the error, the greater the value.

2) MAE

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{y}_t - y_t\right| \tag{18}$$

MAE range is $[0, +\infty)$, the smaller the MAE value is, the higher the prediction accuracy is. When the predicted value is in perfect agreement with the real value, the value is 0. The greater the error, the greater the value.

3) MAPE

$$MAPE = \frac{100\%}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{19}$$

MAPE range is $[0, +\infty)$, similarly, the smaller the value, the higher the prediction accuracy.

The ARIMA-LSTM model proposed in this paper is combined with the IOT data to predict the enterprise economy, and is compared with the original data, the single ARIMA model and the combined model without the IOT data. Figure 8 shows the line chart of the comparison between the predicted value and actual value of enterprise economic forecast by the ARIMA-LSTM combination model combined with IOT data in this paper. As can be seen from Fig. 8, since 2020, due to the impact of COVID-19, the operating revenue of an enterprise has been low for nearly half a year, and gradually approaches the level of previous years in the later period. By June 2021, the operating revenue trend is similar to that of previous years. The difference between the overall predicted value and the real value of ARIMA-LSTM model is small and the error tends to be stable, indicating that ARIMA-LSTM model combined with Internet of Things data can well predict the economic development of enterprises. It shows that ARIMA-LSTM model combined with IOT data can well predict the economic development of enterprises.

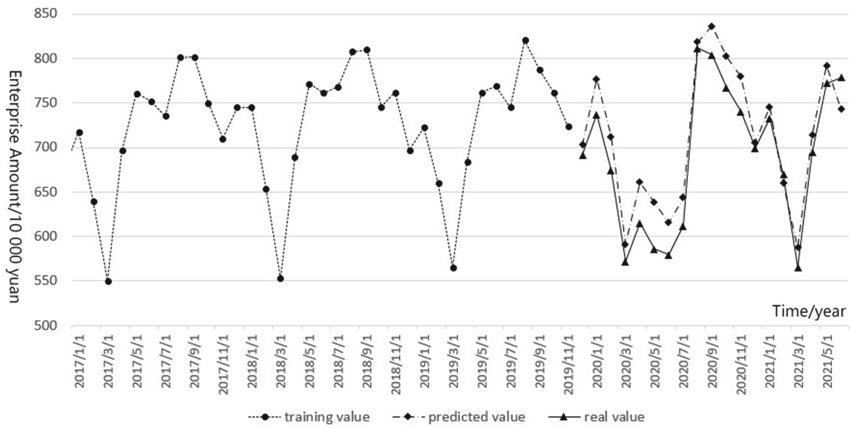


Fig. 8. ARIMA-LSTM enterprise economic forecast results combined with IOT data.

The error and trend prediction between the predicted value and the real value are shown in Table 4. The error is the difference between the predicted value and the true value. When the predicted value is less than the true value, the error is negative. When the predicted value is greater than the true value, the error is positive. The error interval is $[-9.58, 33.39]$, the error percentage interval is $[-1.2\%, 4.5\%]$, and the overall prediction accuracy is higher than 95%. In terms of trend prediction, the combination ARIMA-LSTM model with IOT data can well predict the future trend of enterprise economy and help the government and departments to better control the enterprise economy.

In order to verify the universality of the model, we selected relevant training data input models of several different types of enterprises for training and verification. In order to verify the universality of the model, we select several relevant training data input models of different types of enterprises for training and verification. The prediction results of enterprise A and enterprise B are shown in Fig. 9 and 10 respectively.

The comparison of prediction results between original data, single ARIMA model, combined model without IOT data and the method proposed in this paper is shown in Fig. 11. For easy observation, several types of predictions are shown in the same graph. The solid line is the actual value. The combination model can combine a variety of factors to forecast the economy, which takes more factors into account than the single

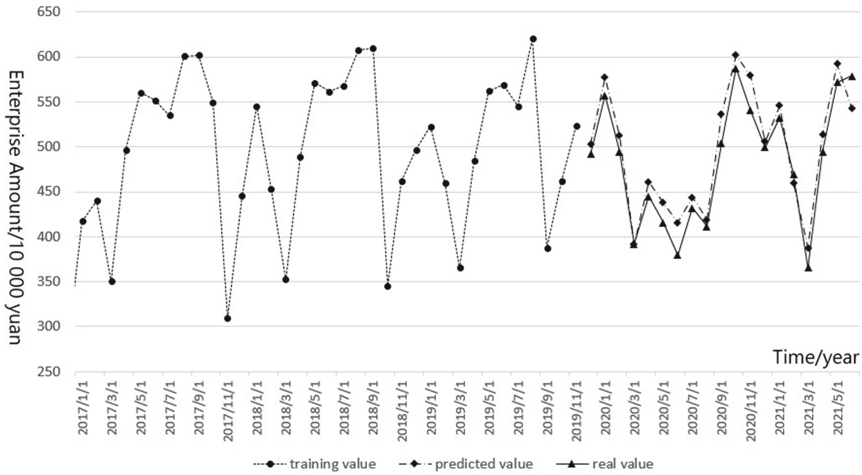


Fig. 9. ARIMA-LSTM economic forecast results of Enterprise A.

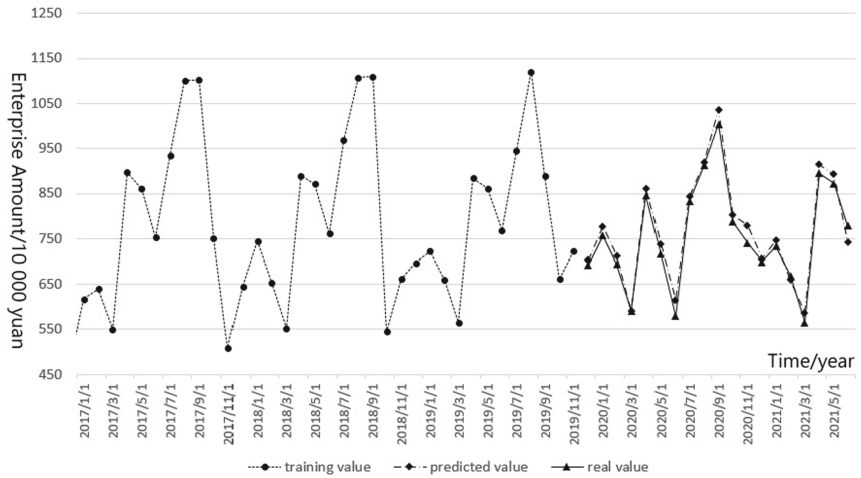


Fig. 10. ARIMA-LSTM economic forecast results of Enterprise B.

model, so the forecast effect is more accurate. In addition, the combined model with IOT data can predict the impact on the enterprise economy in the case of sudden decrease of IOT data caused by shutdown in emergencies (epidemics), because it takes into account the IOT data that can directly reflect the operation status of enterprises. As can be seen from the figure, the original data began to decline from April 2020 and continued until July, when the data did not exceed 650. The prediction curve of the combined model with IOT data is closest to the original data curve. ARIMA and the combined model without IOT data have similar prediction curves, which are closer to the time series data rules of previous years, and cannot predict the economic impact caused by the abnormal operation of enterprises in the early stage of the epidemic. Since 2021, the

Table 4. The prediction results of ARIMA-LSTM combined model.

Time	True value	Predictive value	Error	Trend	
				True	Pridict
2020.1	691.58	703.60	12.02	–	–
2020.2	737.15	770.54	33.39	↑	↑
2020.3	674.53	695.14	20.61	↓	↓
2020.4	571.87	580.27	8.4	↓	↓
2020.5	615.09	637.13	22.04	↑	↑
2020.6	586.00	618.78	32.78	↓	↓
2020.7	579.34	599.48	20.14	↓	↓
2020.8	612.01	635.42	23.41	↑	↑
2020.9	811.59	802.01	−9.58	↑	↑
2020.10	803.82	804.48	0.66	↓	↓
2020.11	767.10	762.93	−4.17	↓	↓
2020.12	740.40	744.18	3.78	↓	↓
2021.1	732.58	745.67	13.09	↓	↓
2021.2	669.64	659.87	−9.77	↓	↓
2021.3	565.34	587.66	22.33	↓	↓
2021.4	694.34	713.99	19.66	↑	↑
2021.5	771.88	792.22	20.34	↑	↑
2021.6	778.79	745.88	−32.91	↑	↓

prediction accuracy of traditional models has gradually declined over time, while the ARIMA-LSTM combination model proposed in this paper, which combines IOT data, can make short-term forecasts based on the actual economic situation of the last month. Smooth errors can be maintained with the actual values in all periods of prediction. The forecast results are more accurate, and more conducive to accurately grasp the direction of future economic development. It can be seen that the predictive value of ARIMA-LSTM model combined with IOT data is the closest to the actual value, and the predictive performance is better than the other three models. It can solve the problems that a single model can't solve, and more accurately predict the direction and value of enterprise economic development. Three evaluation criteria are used to evaluate the two models. The indicators of enterprise economy predicted by different models in the next 16 months are shown in Table 5.

The evaluation results show that the single LSTM model cannot predict the nonlinear relationship with the factors related to economic development, so no comparison is made. The three indexes predicted by single ARIMA model are all larger than those of combination model. Because the single ARIMA model can only predict the time series of future data based on the business revenue data of enterprises, the influencing

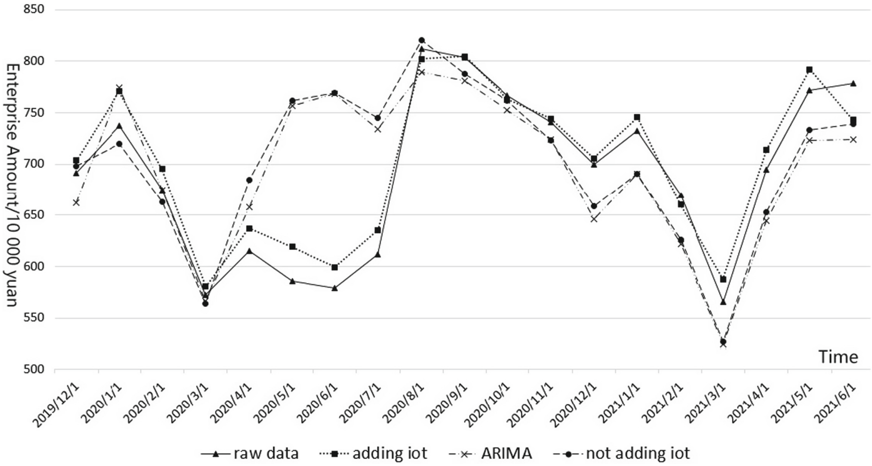


Fig. 11. Comparison of prediction results of different models.

Table 5. Three evaluation indexes of enterprise economy predicted by different models.

Model	RMSE	MAE	MAPE
ARIMA	8683.35	5609.75	9.06%
BP	5663.28	4901.85	8.15%
WNN	5738.10	5012.67	7.63%
The combined model without IOT data	8444.38	5491.72	9.02%
The combined model with IOT data	1913.05	1591.33	2.47%

factor data is single and cannot deal with emergencies. The training data of BP neural network are independent, not affected by other data, and the prediction results are not accurate enough. The number of nodes in the hidden layer of wavelet neural network and the initialization parameters of weights and scale factors between layers are difficult to determine, and the prediction results can not reach the optimal. The three indicators of the combined model without adding IOT data are all medium. It can predict slightly better than a single linear model. However, due to the inability to deal with the impact of emergencies, the prediction results are not accurate enough. The three index values of the combined ARIMA-LSTM model, which combines IOT data, are the smallest compared to other prediction models. It can not only predict the non-linear relationship between a variety of influencing factors on the economic development of enterprises, but also combine the characteristics of the IOT data can reflect the operation of enterprises. It can be timely improved according to the actual situation, making the prediction result more accurate.

5 Conclusion

This paper presents an enterprise economic forecasting method based on ARIMA-LSTM model. ARIMA model is used to predict the influencing factors of the IOT to get the linear predicted value of each column of data. Then LSTM model is used to establish the relationship between the predicted value of IOT data, the influence value of time factor, and the IOT data value of last month, which is important to the forecast result, and predict the final result of enterprise economic data. By comparing the prediction results of other neural network models, it can be seen that ARIMA-LSTM model combined with IOT data solves the problem that a single model cannot predict the impact of multiple influencing factors. At the same time, the addition of IOT data enables the model to predict the impact of emergencies on enterprise economy according to the changes of IOT data. The prediction results are more accurate than other models and combined models without IOT data. Therefore, ARIMA-LSTM model combined with IOT data can more accurately reflect the future economic trend and economic situation of enterprises, which can help government departments better monitor the development prospects of various enterprises and contribute to better decision-making.

References

1. Zhang, C.: Intelligent Internet of things service based on artificial intelligence technology. In: 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), Nanchang, China, pp. 731–734 (2021)
2. Wang, X.: The difference analysis of enterprise operating income and VAT taxable sales. Journal **2019**(09), 25–27 (2019)
3. Xiao, D., Chen, R., Wang, Z.: Performance integration evaluation of investment decision-making of firm based on fuzzy set and BP neural network. Journal **2005**(03), 163–166 (2005)
4. Wang, T.: Research on enterprise economic risk forecast model based on Big Data fusion. In: 2020 International Conference on Big Data and Informatization Education (ICBDIE), Zhangjiajie, China, pp. 5–9 (2020)
5. Barker, J., Gajewar, A., Golyaev, K.: Secure and automated enterprise revenue forecasting. Journal **32**(1), 7657–7664 (2018)
6. Hryhorkiv, V., Buiak, L., Verstiak, A.: Forecasting financial time series using combined ARIMA-ANN algorithm. In: 2020 10th International Conference on Advanced Computer Information Technologies (ACIT), Deggendorf, Germany, pp. 455–458 (2020)
7. Feng, J., Li, H.: Research on macroeconomic forecasting technology based on optimized wavelet neural network. Journal **42**(07), 181–183, 186(2019)
8. Lei, H., Cailan, H.: Comparison of multiple machine learning models based on enterprise revenue forecasting. In: 2021 Asia-Pacific Conference on Communications Technology and Computer Science (ACCTCS), Shenyang, China, pp. 354–359 (2021)
9. Sheng, W., Zhao, H., Sun, Y.: Sales forecasting model based on BP neural network optimized by improved genetic algorithms. Journal **28**(12), 200–204 (2019)
10. Wu, J., Ren, S., Zhang, W.: A daily sales forecasting method based on LSTM model. Journal **30**(2), 133–137 (2020)
11. Xu, M., Ji, C., Zhong, C.: GDP growth forecast method based on electricity consumption. In: Proceedings of 2017 Power Industry Informatization Annual Conference, pp. 224–227. Posts and Telecommunications Press, China (2017)

12. Wang, R., Long, Z.: Research on the relationship between Internet of things and regional economic development. Journal **2020**(12), 147–148 (2020)
13. Chen, Q.: Significance and function of Internet of things in economic development. Journal **2020**(02), 24–25 (2020)
14. Tang, K., Hao, Z., Yixie, B.: Evaluation of prediction methods for parking occupancy rate. Journal **45**(04), 533–543 (2017)
15. Lin, N.: ARIMA model based on ensemble empirical mode decomposition for industry electricity sales prediction. Journal **34**(02), 128–133 (2019)
16. Wang, Z., Zhao, G., Wang, L.: Power consumption forecast of energy-intensive enterprises based on power marketing business data. In: 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), Chengdu, China, pp. 241–245 (2020)
17. Boštjančič, R., Timčenko, V., Kabović, M.: Industrial Internet: architecture, characteristics and implementation challenges. In: 2021 20th International Symposium INFOTEH-JAHORINA (INFOTEH). East Sarajevo, Bosnia and Herzegovina, pp. 1–4 (2021)
18. Akaike, H.: A new look at the statistical model identification. Journal **19**(6), 716–723 (1974)
19. Liu, Y., Li, Q.: Fundamentals of Statistics. Northeast University of Finance and Economics Press, Shenyang, China (2019)