International Journal for Housing Science and Its Applications
Publish August 5, 2025. Volume 46, Issue 3 Pages 4614-4625

https://doi.org/10.70517/ijhsa463395

Dynamic Forecasting Analysis of College Students' Employment and Entrepreneurship Market Trends Based on Long and Short-Term Memory Network Models

Shang Sun[1,*], Di Yang[2] and Juan Hu[3]
[1] School of Economics and Management, Anhui University of Science and Technology, Huainan, Anhui, 232001, China
[2] School of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan, Anhui, 232001, China
[3] Huainan Vocational and Technical College, Huainan, Anhui, 232001, China
Corresponding authors: (e-mail: sunshang24@163.com).

Abstract This paper first establishes a hierarchical employment and entrepreneurship quality coordinated development analysis system, through the use of time series ARIMA model and neural network LSTM model, respectively, for short-term prediction of college students' employment and entrepreneurship market dynamics trends. In view of the prediction defects of a single model, this paper adopts the optimal weighted combination prediction method to optimize the combination of two single models, obtains the ARIMA-LSTM combination measurement model, and analyzes the application effect of the three models with actual cases. The results show that the relative errors of single models ARIMA and LSTM in predicting the employment and entrepreneurship of college students in the validation set are 9.52% and 10.13% respectively, and the R^2 is 83.0378 and 81.2749, which shows that the predictive effect of the models is general. The ARIMA-LSTM combination model is significantly better than the single model ARIMA and LSTM in terms of the accuracy and stability of the prediction of college students' employment and entrepreneurship, at this time, the correlation indexes of the three models of ARIMA, LSTM and ARIMA-LSTM are 0.8064, 0.7959 and 0.9773, respectively, which can be seen that the combination model can effectively integrate the two single model s linear forecasting ability and nonlinear time series modeling advantages, thus improving the accuracy and reliability of the ARIMA-LSTM model in predicting the dynamics of college students' employment and entrepreneurship market trends.

Index Terms ARIMA model, LSTM model, ARIMA-LSTM model, Employment and entrepreneurship market trend

I. Introduction

With the economic development and technological progress, the situation of college students' employment and entrepreneurship is also changing [1], [2]. In this era, influenced by the Internet and digitalization, college students should have excellent technological literacy, master and apply emerging technologies, and seek opportunities and breakthroughs in employment and entrepreneurship, and a reasonable prediction of the dynamics of employment and entrepreneurship market trends in this context can provide more references for college students' employment and entrepreneurship [3]-[6].

In recent years, the situation of the employment market has undergone some unusual changes. On the one hand, the number of college graduates is increasing year by year, and the competition is fiercer than ever [7]. On the other hand, the demand for talents in many traditional industries is decreasing, while emerging fields need talents with brand new skills [8]. Therefore, the employment situation of college students is facing greater challenges. In this situation, college students not only need to have excellent skills and knowledge in specialized fields, but also need to have cross-border and comprehensive qualities [9]-[11]. At the same time, the government and enterprises also need to clarify their employment needs and strengthen the employment policy and guidance for college students [12].

In the process of constantly adapting to change, college students can better realize the goals of employment and entrepreneurship. With the emergence of many new fields in the past few years, it brings more opportunities for college students' employment and entrepreneurship [13]. Among them, the most notable one is the Internet industry. From Internet entrepreneurship, data analysis, online media, e-commerce technology and mobile applications, to innovative technologies such as artificial intelligence, virtual reality, blockchain and drones, all of them provide brand new employment and entrepreneurship opportunities for college students [14]-[17]. In addition, new agriculture, new manufacturing, new logistics and new energy industries have emerged, many of which require a great deal of expertise and practical experience [18], [19]. These above development trends have important reference value for the fledgling college students.

This paper first establishes a hierarchical employment quality coordinated development indicator analysis system containing five primary indicators, eleven secondary indicators, and twenty-seven tertiary indicators. After that, the ARIMA time series model and the LSTM neural network model are introduced respectively, and the simulation prediction effect of the two single models on the dynamic trend of college students' employment and entrepreneurship market is analyzed. Subsequently, this paper also combines and optimizes the ARIMA model and LSTM neural network model through the optimal weighted combination prediction method, and obtains the combination measurement model ARIMA-LSTM based on the college students' employment and entrepreneurship index, as well as the principle of its construction and calculation formula. The root mean square error is explicitly used in the experiment to compare the simulated valuation effects of the combination model and each single model on the elements of college students' employment and entrepreneurship, and to test the reasonableness of constructing the ARIMA-LSTM combination model, so as to accurately predict the dynamic trend of the college students' employment and entrepreneurship market.

Table 1: The selection results of specific employment quality indexes

<table>
<tr><th>Primary indicator</th><th>Secondary indicator</th><th>Tertiary index</th><th>Index direction</th></tr>
<tr><td rowspan="5">The environment of employment and entrepreneurship</td><td>Development and employment</td><td>Employment elasticity</td><td>Positive</td></tr>
<tr><td rowspan="3">Employment fairness</td><td>Urban and rural income gap</td><td>Negative</td></tr>
<tr><td>Income gap of individual ownership of urban units</td><td>Negative</td></tr>
<tr><td>Proportion of women in urban and urban units</td><td>Positive</td></tr>
<tr><td>Labor market segmentation</td><td>The proportion of local hukou to the local hukou</td><td>Positive</td></tr>
<tr><td rowspan="6">The ability to work and start a business</td><td rowspan="3">Education level</td><td>The average person is affected by education</td><td>Positive</td></tr>
<tr><td>College employment ratio</td><td>Positive</td></tr>
<tr><td>Undergraduate and above employment</td><td>Positive</td></tr>
<tr><td rowspan="3">Vocational skill</td><td>Primary vocational certificate acquisition rate</td><td>Positive</td></tr>
<tr><td>Rate of classification of intermediate vocational certificate</td><td>Positive</td></tr>
<tr><td>Professional certificate acquisition rate for technicians, advanced, senior technicians</td><td>Positive</td></tr>
<tr><td rowspan="9">Remuneration of labor</td><td rowspan="4">Wage level</td><td>The average disposable wage income of residents</td><td>Positive</td></tr>
<tr><td>Average wage growth</td><td>Positive</td></tr>
<tr><td>Minimum wage</td><td>Positive</td></tr>
<tr><td>The total amount of wages is the proportion of GDP</td><td>Positive</td></tr>
<tr><td rowspan="5">Social security</td><td>Participation rate of urban endowment insurance</td><td>Positive</td></tr>
<tr><td>Medical insurance participation rate</td><td>Positive</td></tr>
<tr><td>Participation rate of inductrial injury insurance</td><td>Positive</td></tr>
<tr><td>Unemployment insurance participation rate</td><td>Positive</td></tr>
<tr><td>Fertility insurance participation rate</td><td>Positive</td></tr>
<tr><td rowspan="5">Labor protection</td><td rowspan="3">Labor relation</td><td>The ratio of the case to the case of arbitration</td><td>Negative</td></tr>
<tr><td>Settlement ratio</td><td>Positive</td></tr>
<tr><td>The membership of the union members occupies the employment ratio of the urban units</td><td>Positive</td></tr>
<tr><td rowspan="2">Labor safety</td><td>Occupational disease</td><td>Negative</td></tr>
<tr><td>Employee injury rate</td><td>Negative</td></tr>
<tr><td rowspan="2">Government service</td><td>Employment service</td><td>Occupational guidance</td><td>Positive</td></tr>
<tr><td>Business service</td><td>The number of entrepreneurial services</td><td>Positive</td></tr>
</table>

II. Construction of ARIMA and LSTM models based on university students' employment and entrepreneurship

II. A. Indicator system of employment and entrepreneurship for university students

For the measurement of employment quality, this paper, based on the availability of data, the relevance and simplicity of the index system, and the combination of previous research, screens and identifies the indices that most directly reflect the quality of employment, and establishes the employment quality index system. The employment quality index system established in this paper consists of five major first-level indicators, which are: employment environment, employability, labor remuneration, labor protection, and government services.

The first-level indicators of employment environment include three second-level indicators of economic development and employment, employment equity, and labor market segmentation; employability is divided into two second-level indicators of education level and vocational skills; labor remuneration is measured comprehensively by two second-level indicators of wage level and social security status; labor protection includes two second-level indicators of labor relations and labor safety; and government services are embodied in the aspects of employment services and entrepreneurship services.

Accordingly, this paper establishes an employment quality evaluation system with five first-level indicators, eleven second-level indicators and twenty-seven third-level indicators, and the results of the selection of specific employment quality indicators are shown in Table 1.

II. B.ARIMA prediction model and theory

II. B. 1) Analysis and definition of the ARIMA model

The ARIMA model [20] consists of three components, namely autoregressive (AR), difference (I) and moving average (MA), which have the following meanings: AR denotes autoregressive; I denotes a single integer; and MA denotes a moving average model.

(1) AR model

AR model means that the value of the current time series is only determined by the value of the past time series. The mathematical expression is as follows:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \cdots + \alpha_p y_{t-p} + \varepsilon_t \tag{1}$$

where α_0 denotes the constant term, $\alpha_1, \alpha_2 \cdots \alpha_p$ denotes the coefficients of the autoregressive terms of each order, ε_t denotes the sequence of white noises that are independent of each other, y_t denotes the current predicted value, and $y_{t-1}, y_{t-2}, \cdots y_{t-p}$ denotes the time series values at past moments. From the above equation, it can be seen that the AR model is determined only by the previous time series values and residuals. The autoregressive model is one of the simpler models in the time series model, and is used relatively frequently in practical applications, and is often used as a comparison model.

(2) MA model

MA model has nothing to do with the previous time series value, it is a linear model to predict the current value by using a linear combination of the weighted average of the current ε_t and the random errors $\{\varepsilon_t\}$ of the past q moments, the mathematical expression can be expressed as:

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \tag{2}$$

where q is said to be the order moving average model, denoted as the $MA(q)$ model, where $\theta_1, \theta_2 \cdots \theta_q$ are the moving average coefficients, and $\{\varepsilon_t\}$ is the sequence of random disturbing terms in different periods. When $\theta_0 = 0$, it is called the $MA(q)$ model.

(3) ARMA model

The ARMA model is a new model obtained by recombining the AR model and the MA model. This model is a linear model in which previous time series values and residual series are combined. The mathematical expression can be expressed as:

$$y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} \ldots + \alpha_p y_{t-p} + \varepsilon_t - \beta_1 \varepsilon_{t-1} - \beta_2 \varepsilon_{t-2} \ldots - \beta_q \varepsilon_{t-q} \tag{3}$$

p is the autoregressive order and q is the sliding average order, which is abbreviated as the $ARMA(p,q)$ model. Where $\alpha_0, \alpha_1, \alpha_2, \ldots \alpha_p$ and $\beta_1, \beta_2, \ldots, \beta_q$ are parameters.

(4) ARIMA model

AR, MA, ARMA models are all applicable to smooth stochastic processes, but most of the time series obtained in our practical work are non-smooth series. In order to solve the non-stationary series thus appear ARIMA model, ARIMA model can be regarded as the ARMA model after differencing, the mathematical expression of the model is:

$$\Phi(B) \nabla^d y_t = \Theta(B) \varepsilon_t \tag{4}$$

where d is the difference order, $\Phi(B)=1-\phi_1 B-\cdots-\phi_p B^p$ denotes the autoregressive coefficient polynomial, and $\Theta(B)=1-\theta_1 B-\cdots-\theta_q B^q$ denotes the moving smoothing coefficient polynomial.

II. B. 2) Construction of the ARIMA model

The construction steps of ARIMA model can be divided into several steps: time series smoothness detection, non-smooth series processing, pattern recognition, model optimization and model testing:

(1) Smoothness detection of time series

The detection of smoothness is to determine whether the current time series in the subsequent period of time to the existing trend continues, if the current time series in the future period of time to have this trend, then its research and analysis is meaningful. Smooth time series is the mean, variance, covariance and other statistical properties are constants; smoothness of the test methods are mainly graphical test method and ADF single root test.

Graphical test is to observe the plotting of the time series of the time series, if the time series is around a mean value of random fluctuations up and down, it proves that the series is a smooth series, if you find that the changes in the series of curves without a fixed trend, the mean, the variance of the change, then it can be concluded that the series is non-smooth series.

ADF single-root test is the first assumption that the sequence is the existence of a unit root, if the test statistic value obtained through the ADF test is much less than 1% of the corresponding critical value, then it can be shown that the time series is smooth.

(2) Non-stationary series processing

Through the time series of the smoothness of the test, if the time series is non-smooth, then the difference operation.

(3) Pure randomness test

After the smoothness test, the time series also need to carry out a pure randomness test (also known as white noise test), is to check the time series to determine whether there is no correlation between the time series, the historical value of the time series has an impact on the future value. If the past historical value of the time series does not have the slightest impact on the future development trend, it means that the time series has no research value, and the construction of the model can be terminated.

(4) Model Identification

Pattern recognition is the process of choosing a model that better fits the trend of the time series from among the AR, MA, and ARMA models, the specific process is to use the nature of the autocorrelation coefficients and partial autocorrelation coefficients to determine the autoregressive model should be chosen if the partial autocorrelation function is truncated at p while the autocorrelation function tends to be zero $AR(p)$; if the autocorrelation function is truncated at q, and the value of the bias correlation function tends to zero, then the mobile regression $MA(q)$ model should be selected; if neither the autocorrelation nor the bias correlation function is truncated and the values of the functions tend to zero, then the $ARMA(p,q)$ model should be selected.

(5) Model Optimization

Model optimization is generally through the comparison of AIC value and BIC value, and then select the optimal model.

In general, AIC is defined as:

$$AIC = 2k - 2\ln(L) \tag{5}$$

where k is the number of model parameters and L is the likelihood function.

In general, the BIC is defined as:

$$BIC = k\ln(n) - 2\ln(L) \tag{6}$$

where K and L have the same meaning as AIC and n is the sample size.

The AIC criterion is used to obtain the optimal parameters of the model by finding a balance between the complexity of the model and the likelihood function, while the expression of BIC increases the number of samples on top of AIC, which prevents the case when the number of samples is too large in order to improve the accuracy, thus increasing the complexity of the model. So it is said that the BIC criterion is more suitable for the case of large sample capacity.

(6) Model Testing

Model testing is to carry out significance testing and parameter testing of the model.

The significance test of the model is to check whether the model can fully explain the correlation of the time series, if the residual series of the time series is white noise series, it passes the test, otherwise it needs to re-fit the model, and then validate the model.

The model parameter test is to make the model most streamlined by removing those dependent variables that do not have a significant impact on the results.

(7) Model Prediction

Input the time interval to be predicted for prediction analysis.The ARIMA model construction process is shown in Figure 1.

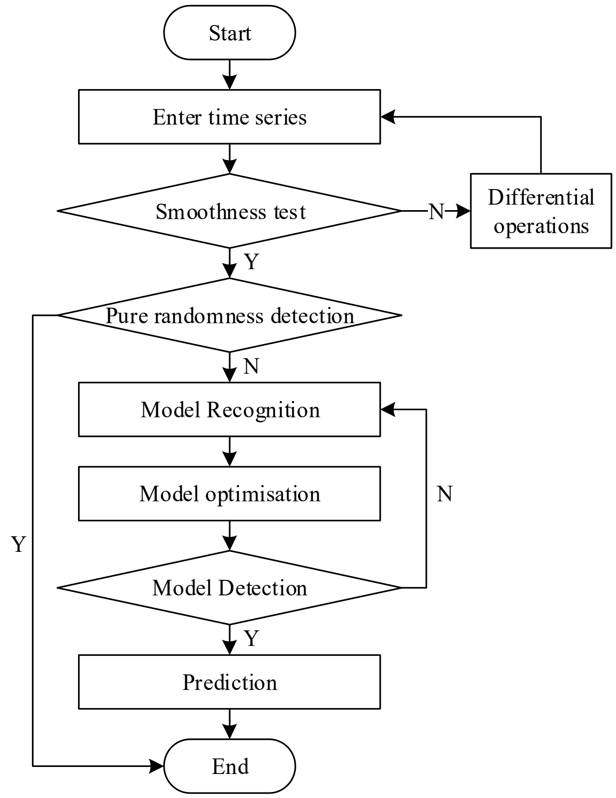


Figure 1: The arima model builds the flowchart

II. C.LSTM Neural Network

The Long Short-Term Memory (LSTM) neural network model [21] is a variant based on Recurrent Neural Networks (RNN). While RNNs can only have short-term memory due to gradient vanishing, LSTM solves the gradient vanishing problem by introducing well-defined memory units and gate structures.

The following are the principles related to long and short term memory neural networks:

There are three inputs to the implicit layer of the LSTM neural network, namely the unit state c_{t-1} at the previous moment, the output value s_t at the previous moment, and the current input value x_t .

The management control of the cell state c is the key for LSTM neural network to be able to remember the information for a long period of time, and LSTM realizes this control by introducing three gate structures: forgetting gate, input gate, and output gate.

(1) The forgetting gate is used to control the saving of information to the current unit state c_t that needs to be saved in the unit state c_{t-1} of the previous moment, the expression is:

$$f_t = \sigma\left(W_f \cdot [s_{t-1}, x_t] + b_f\right) \tag{7}$$

where W_f is the weight matrix of the forgetting gate, σ is the sigmoid activation function of the forgetting door, b_f is the offset of the forgetting door, x_t is the input of the current moment, s_{t-1} is the unit state of the previous moment, and the final f_t is the value of a $[0,1]$, if $f_t=1$ means that all the unit states of the previous moment are remembered, and if $f_t=0$ The value of f_t is generally $(0,1)$, and only the information that needs to be saved in the unit state at the previous time is remembered.

(2) The input gate controls how much information about the input x_t of the current moment is stored in the current cell state c_t, with the expression:

$$i_t=\sigma\left(W_i\cdot[s_{t-1},x_t]+b_i\right) \tag{8}$$

where W_i is the weight matrix of the forgetting gate, σ is the sigmoid activation function of the forgetting gate, b_i is the bias of the forgetting gate, x_t is the input of the current moment, s_{t-1} is the state of the unit in the previous moment, and the last obtained i_t is a value of $[0,1]$, if $i_t=1$ indicates that all the inputs at this time are memorized, on the contrary, if $i_t=0$, it indicates that all the inputs at this time are forgotten, and the value of i_t is generally $(0,1)$, which is only the information that needs to be saved in the inputs at this time is memorized.

(3) Getting the candidate value vector $\tilde{c}_t$ of the current unit state is the prerequisite to get the current unit state c_t, and the expression of $\tilde{c}_t$ is:

$$\tilde{c}_t=\tanh\left(W_C\cdot[s_{t-1},x_t]+b_C\right) \tag{9}$$

where W_C is the weight matrix of the oblivious gate, $\tanh$ is the activation function of the oblivious gate, b_C is the bias of the oblivious gate, x_t is the input at the current moment, s_{t-1} is the cell state at the previous moment, and finally, $\tilde{c}_t$ is obtained as a value of $[0,1]$.

(4) The expression for the current cell state c_t is as follows:

$$c_t=f_t*c_{t-1}+i_t*\tilde{c}_t \tag{10}$$

where c_{t-1} is the last moment unit state, $\tilde{c}_t$, i_t, f_t is the value of the current moment, and all can be obtained by the above formula, you can integrate the current memory and long-term memory and together to get the current unit state c_t.

(5) The output gate controls the amount of information saved in the current output s_t in the current unit state c_t, the expression is shown below:

$$o_t=\sigma\left(W_o\cdot[s_{t-1},x_t]+b_o\right) \tag{11}$$

$$s_t=o_t*\tanh(c_t) \tag{12}$$

where W_o is the weight matrix of the oblivion gate, σ is the sigmoid activation function of the oblivion gate, b_o is the bias of the oblivion gate, x_t is the input of the current moment, s_{t-1} is the state of the cell in the previous moment, and the final o_t obtained is a $[0,1]$ value, and finally, the current moment is integrated with the output gate to obtain the output s_t. cell is integrated with the o_t obtained from the output gate to get the current moment output s_t.

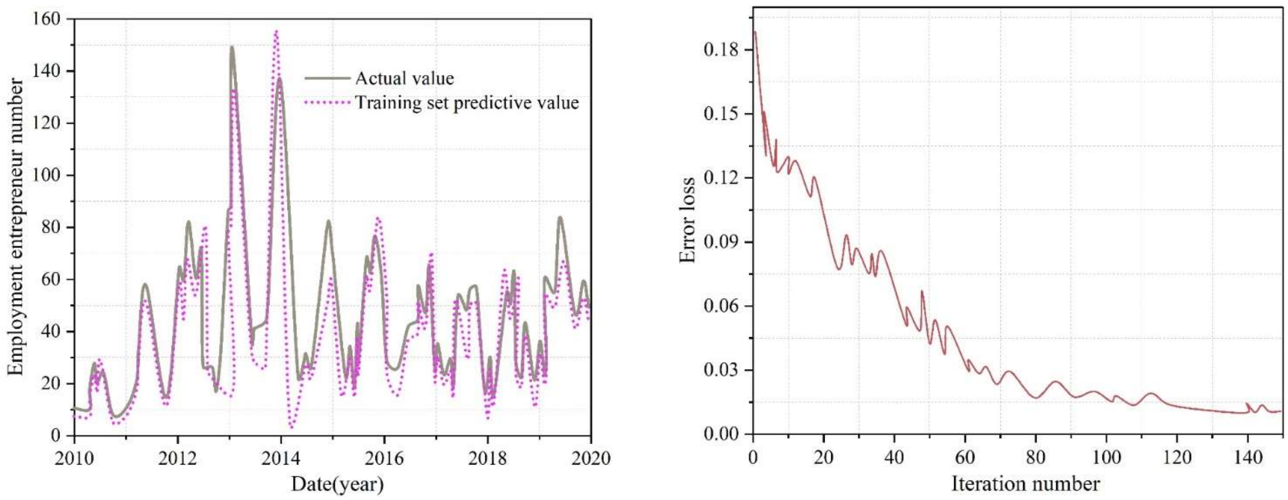
II. D.ARIMA and LSTM models for predicting college students' employment and entrepreneurship

II. D. 1) Data sources and processing

First build a PyTorch environment, and then normalize the time series data, using the MinMaxScaler class in the sklearn.preprocessing module to scale the data to within the $[-1,1]$ value range. The purpose of the scaling is to narrow down the influence weights of each feature, as well as to avoid the problem of gradient disappearance and gradient explosion for long time training, and to improve the accuracy. And the input and output columns of the training and validation sets are divided into x and y, and the input columns are converted into three-dimensional arrays. The input data format should be converted to [samples, timesteps, features], samples denotes the number of data, timesteps denotes the input step size, features denotes the number of features of the data.

II. D. 2) Analysis of the prediction results of the LSTM model

An LSTM layer is created using nn.Module in PyTorch and the loss function is trained. The LSTM model training fitting results and loss values are shown in Fig. 2, where (a) and (b) represent the training fitting results and loss values respectively. The model is trained using the training set and validated using the validation set. The root mean square error (MSE) is used as the loss function to train for 150 iterations. The optimizer function is chosen to be the adam optimizer. It can be seen that the root mean square error of the loss function is gradually decreasing during the training process.



(a)Training fitting results (b)Loss value

Figure 2: LSTM model training fitting results and loss values

The trained LSTM model is used to predict and evaluate the indicators on the data in the validation set, and the results are inverse normalized to reduce the data to the range of values of the original data, and Figure 3 shows the results of the LSTM model for predicting the time series. It can be found that the LSTM model has a large error in predicting the time series results.

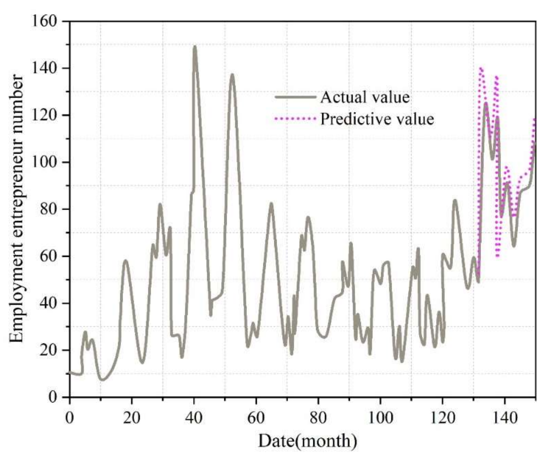


Figure 3: The LSTM model predicts the timing results

II. D. 3) Analysis of prediction results of ARIMA and LSTM models

The constructed ARIMA and LSTM models were used to predict the employment and entrepreneurship of 776 college students from five highly efficient universities in M province in 2023-2024, and the number of employment and entrepreneurship were 452 and 324, respectively. The fitting and prediction effects of ARIMA and LSTM models were evaluated using the indicators RMSE, MAE and R^2.The results of the comparison of the prediction of ARIMA and LSTM models are shown in Table 2; the results of the comparison of the evaluation of ARIMA and LSTM models are shown in Table 3. The ARIMA model and LSTM model are used to predict the monthly employment number of college students' employment and entrepreneurship in the validation set, and compared with the actual value. From the prediction effect, the relative error of the validation set predicted by the LSTM model in the number of college students' employment and entrepreneurship employment is 10.13%, which is larger than the relative error of the ARIMA model of 9.52%, which indicates that the prediction effect of the ARIMA model is better, and all the indexes in the training set and validation set of the ARIMA model are better than those of the LSTM model.

Table 2: The ARIMA and LSTM models predict the comparison results

Date	Actual number	ARIMA			LSTM		
		Forecast population	Absolute error	Relative error	Forecast population	Absolute error	Relative error
2021.9	46	57	-11	7.07	51	-5	7.84
2021.10	19	13	2	40.83	30	-11	108.77
2021.11	25	16	9	23.86	27	-2	7.73
2021.12	21	14	7	28.35	24	-3	29.56
2022.1	43	37	6	13.06	23	20	100.45
2022.2	46	45	1	30.8	40	6	38.95
2022.3	68	75	-7	22.69	62	6	39.67
2022.4	93	97	-4	34.63	98	-5	9.37
2022.5	103	89	14	79.65	108	-5	26.47
2022.6	98	74	24	206.48	94	4	5.11
2022.7	102	137	-35	267.16	73	29	151.88
2022.8	37	49	-12	45.22	51	-14	35.3
2022.9	34	26	8	4.38	30	4	5.05
2022.10	11	7	4	24.51	30	-19	131.12
2022.11	11	17	-6	22.36	27	-16	75.7
2022.12	19	15	4	13.48	24	-5	17.84
Tot	776	768	4	9.52	792	-16	10.13

Table 3: ARIMA and LSTM model evaluation comparison results

Model	Training set			Verification set	
	RMSE	MAE	R^2	RMSE	MAE
LSTM	20.6402	18.494	81.2749	15.2018	9.6286
ARIMA	16.2712	13.8231	83.0378	13.7565	8.2548

III. Construction of ARIMA-LSTM combination model and application effect

III. A. Constructing the combined ARIMA-LSTM model

Based on the confirmation that the two single models, ARIMA time series and LSTM neural network, can effectively simulate college students' employment and entrepreneurship data, in order to further improve the simulation accuracy of the relevant college students' employment and entrepreneurship data and reduce the prediction error of the dynamic trend of the college students' employment and entrepreneurship market, this paper adopts the optimal weighted combination forecasting method to optimize the combination of ARIMA and LSTM models, so as to obtain a more reliable and objective combination model. According to the simulation results of ARIMA and LSTM single models, and the actual values of relevant college students' employment and entrepreneurship data, this paper adopts the optimal weighted combination prediction method to optimize the combination of ARIMA and LSTM neural network models, so as to obtain a more reliable and objective combination model, and to effectively improve the simulation accuracy of college students' employment and entrepreneurship factors.

The purpose of constructing ARIMA-LSTM combined model [22], [23] is to improve the prediction accuracy of college students' employment and entrepreneurship simulation as much as possible by comprehensively utilizing the information provided by each single model, and theoretically, this combined model can realize a more comprehensive and systematic consideration of the problem than a single model.

The method of determining the weighting coefficients of each single model in the ARIMA-LSTM combined prediction model constructed in this paper: based on the combined model of college students' employment and entrepreneurship data simulation and prediction results of the minimum error sum of squares criterion, the weighting coefficients of each single model are derived through the form of linear equations. The resulting combined measurement model using ARIMA model and LSTM neural network model better simulates and predicts the dynamic change trend of college students' employment and entrepreneurship market, and improves the simulation and prediction accuracy by comprehensively utilizing the information provided by multiple single models.

Constructing the ARIMA-LSTM combination model

Let A and L be the simulated prediction values of college students' employment and entrepreneurship elements by ARIMA model and LSTM neural network model respectively, and Y be the prediction value of the optimal combination, and their simulated prediction errors are E_a, E_1, and E_y, respectively. According to the principle that the minimum of the sum of squares of prediction errors is the optimal principle, i.e., letting $E=\sum E_y^2$ to take the minimum of is optimal, K_a and K_l are the corresponding weight coefficients $(K_a+K_1=1)$, then we have the following formula for E_y and Y:

$$E_y = K_a * E_a + K_1 * E_1 \tag{13}$$

$$Y = K_a A + K_1 L \tag{14}$$

In the above equation, the equations for K_l and K_a are calculated as:

$$K_l = 1 - K_a \tag{15}$$

$$K_a = -\frac{\sum (E_a - E_1)E_1}{\sum (E_a - E_1)^2} \tag{16}$$

$$RMSE = \left(\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2\right)^{\frac{1}{2}} \tag{17}$$

By determining the weight ratios of two single college student employment and entrepreneurship factor measurement models, ARIMA time series and LSTM neural network, in the optimal combination model, the simulated prediction values of the ARIMA-LSTM college student employment and entrepreneurship assessment index combination measurement model on the time series of the dynamic trend of change in the employment and entrepreneurship market are weighted.

In order to further quantitatively compare the valuation results of the ARIMA-LSTM combination model with those of each single model, this paper uses the root mean square error (RMSE) to test and compare the valuation effects of the ARIMA-LSTM combination model and the two single models, in order to validate the reasonableness of the combination model constructed in this paper, and to reduce the prediction errors of the comprehensive index of assessment of employment and entrepreneurship of college students.

III. B. Analysis of the prediction effect of ARIMA-LSTM models

In order to further assess the advantages and disadvantages of ARIMA model, LSTM model and ARIMA-LSTM combination model in the prediction of college students' employment and entrepreneurship, the prediction results of the combination model and the results of the error analysis of the three models are presented in Figures 4 and 5, respectively. From the figures, it can be observed that under the same conditions, compared with the ARIMA model and LSTM model, the volatility of the combined ARIMA-LSTM model in time series forecasting is very obviously smaller than the other two models. This result suggests that the combined model may have better stability and forecasting accuracy in dealing with time series data.

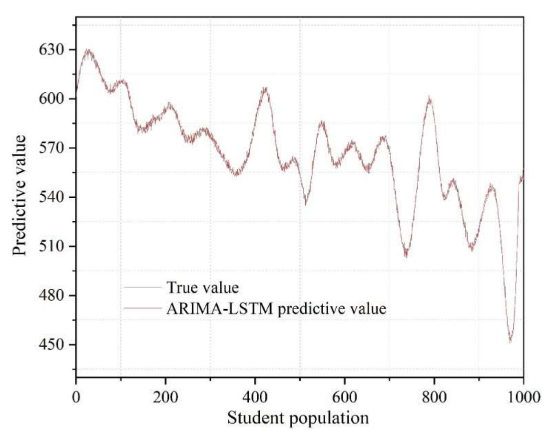


Figure 4: The prediction of the composite model

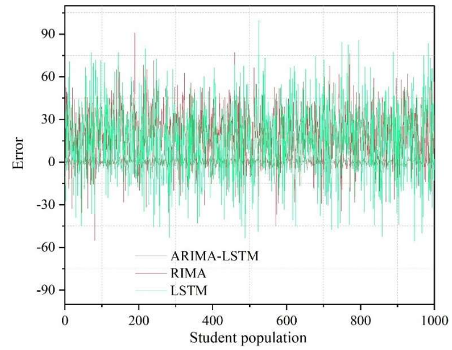


Figure 5: The results of the error analysis of the three models

To further verify whether the performance of the combined model exceeds that of the single model, the performance of the three models on their respective key evaluation metrics is compared in detail. The results of the evaluation indexes of different models are shown in Table 4. From the analysis of the data in the table, it can be seen that the combined ARIMA-LSTM model is highly significantly better than the separate ARIMA and LSTM models in terms of both mean error and confidence interval, while the difference in correlation coefficients among the three models is small. This result indicates that although the performance of the three models in capturing the correlation of the time series data also differs greatly, the combined model performs better in reducing the prediction error, and the accuracy of the correlation coefficient of the ARIMA-LSTM model reaches 97.73%, and the effect is closer to the real value in terms of the stability and accuracy of the prediction.

Table 4: Results of evaluation of different models

Model name	Error mean	Confidence interval	Correlation coefficient
ARIMA	5.2728	(5.2,8.5)	0.8064
LSTM	-8.3044	(30.4, -18.3)	0.7959
ARIMA-LSTM	3.0001	(-0.5, 0.5)	0.9773

In order to verify the generalization ability of the model, taking Henan Province as an example, the model was used to predict 4741 students in 7 schools in the province, and the prediction results of college students' employment and entrepreneurship in some universities in Henan Province are shown in Table 5; the results of evaluation indexes of different models are shown in Table 6. It can be seen that the prediction effect of the combined model in college students' employment and entrepreneurship in other provinces is better than that of the ARIMA model and the LSTM model, especially the correlation index of the combined model (0.9908) is higher than that of the single models of ARIMA and LSTM by 0.2187 and 0.2537, respectively.

Considering the evaluation indexes together, it can be concluded that the combined ARIMA-LSTM model is significantly better than the single ARIMA and LSTM models in terms of accuracy and stability in predicting college students' employment and entrepreneurship. This combined model effectively integrates the linear forecasting ability of the ARIMA model and the nonlinear time series modeling advantage of the LSTM model, thus improving the accuracy and reliability of forecasting. As a result, the combined ARIMA-LSTM model exhibits better performance when dealing with complex time series forecasting problems and provides a more effective method for forecasting.

Table 5: The results of students' employment entrepreneurship

School name	True value	Predictive value	Error
Beijing University	700	699	1
Nanjing university	670	671	-1
Zhejiang university	681	681	0
Zhongshan university	656	653	3
Shanghai jiaotong university	689	689	0
Beijing university of aeronautics and astronautics	668	669	-1
science and technology university of China	677	675	2

Table 6: Different model evaluation results

Model name	Error mean	Confidence interval	Correlation coefficient
LSTM	-21.5555	(-27.3,13.7)	0.7371
ARIMA	19.4026	(-12.9,24.6)	0.7721
ARIMA-LSTM	-0.3489	(-0.3,0.1)	0.9908

IV. Conclusion

In this study, we first constructed the evaluation index system of college students' employment and entrepreneurship, proposed a combined measurement model of college students' employment and entrepreneurship index of ARIMA-LSTM, and compared and analyzed the prediction results of the single model and combined model of ARIMA and LSTM.

(1) The effect of ARIMA model fitting and prediction is better than LSTM model, but the relative error of both models in predicting the dynamic trend of college students' employment and entrepreneurship market is larger, especially the relative error of LSTM model prediction is more than 10%, which shows that the effect of two single models in college students' employment and entrepreneurship prediction is general.

(2) The experimental results in the application of college students' employment and entrepreneurship prediction show that the combined model has extremely significant advantages in improving the prediction accuracy and model stability, and at this time, the correlation coefficient accuracy of the ARIMA-LSTM model reaches 97.73%, which provides an effective solution for college students' employment and entrepreneurship, as well as for other time series prediction tasks.

Funding

1. General Project of Humanities and Social Sciences Research of Ministry of Education in 2023 (23JDSZ3081).

2. Key research project of Anhui University of Science and Technology's school-level teaching reform in 2023 (JG202344).

References

[1] Zhu, R., Zhao, G., Long, Z., Huang, Y., & Huang, Z. (2022). Entrepreneurship or employment? A survey of college students' sustainable entrepreneurial intentions. Sustainability, 14(9), 5466.

[2] Yin, L. (2022). From employment pressure to entrepreneurial motivation: An empirical analysis of college students in 14 universities in China. Frontiers in psychology, 13, 924302.

[3] Fu, H. (2022). Evaluation model of employment and entrepreneurship of university students based on classification algorithm. Mathematical Problems in Engineering, 2022(1), 2477264.

[4] Waldemar Staniewski, M., & Szopiński, T. (2015). Student readiness to start their own business. Economic research-Ekonomska istraživanja, 28(1), 608-619.

[5] Portuguez Castro, M., & Gómez Zermeño, M. G. (2021). Identifying entrepreneurial interest and skills among university students. Sustainability, 13(13), 6995.

[6] Sun, S., & Di Yang, J. H. (2025). Dynamic Analysis of College Students' Employment and Entrepreneurship Market Trends Based on Time Series Forecasting Models. J. COMBIN. MATH. COMBIN. COMPUT, 127, 65-82.

[7] Jiang, L., Li, Y., & Jiang, C. (2017). Employment competitiveness of college students based on improved AHP. Eurasia Journal of Mathematics, Science and Technology Education, 13(8), 5225-5232.

[8] Chen, J., Wu, J., & Jiang, W. (2021, March). Demand analysis of employment talents based on deep learning. In IOP Conference Series: Earth and Environmental Science (Vol. 692, No. 4, p. 042017). IOP Publishing.

[9] Yaghi, A., & Alabed, N. (2021). Career decision-making difficulties among university students: does employment status matter?. Higher Education, Skills and Work-Based Learning, 11(5), 1143-1159.

[10] Mokibelo, E., & Seru, L. (2020). Learning for Employment: Challenges Encountered by Students. Open Access Library Journal, 7(7), 1-16.

[11] Zhongyuan, L. (2022). Employment Analysis of College Students Under the Current Economic Situation. International Journal of Management and Education in Human Development, 2(01), 296-302.

[12] Huang, L. (2022). The establishment of college student employment guidance system integrating artificial intelligence and civic education. Mathematical Problems in Engineering, 2022(1), 3934381.

[13] Fasla, N. P. (2017). A STUDY ON ENTREPRENEURSHIP ATTITUDE AMONG COLLEGE STUDENTS. International Journal of Research Science and Management, 4(4), 53-59.

[14] Li, D. (2018, September). A brief analysis of college students using the internet to start their own business. In 2018 4th International Conference on Social Science and Higher Education (ICSSHE 2018). Atlantis Press.

[15] Chen, S., Zhang, Q., Zhao, Q., Deng, H., & Su, Y. S. (2021). Influencing factors of university relocation on college students' intention to engage in local entrepreneurship and employment. Frontiers in Psychology, 12, 750972.

[16] Otache, I. (2019). Entrepreneurship education and undergraduate students' self-and paid-employment intentions: A conceptual framework. Education+ training, 61(1), 46-64.

[17] Presley, A., Presley, T., & Blum, M. (2018). Sustainability and company attractiveness: A study of American college students entering the job market. Sustainability Accounting, Management and Policy Journal, 9(4), 470-489.

[18] Li, Y. (2022). What Attracts College Students to Start a Business in Rural Areas?—A Qualitative Comparative Analysis from China. Sustainability, 14(22), 15459.
[19] Asgarova, V., Kan, Y., Reindl, S., & Liu, M. (2023). What changes are university students seeking? Bringing employers to classrooms to facilitate post-graduation opportunities. Irish Educational Studies, 42(4), 825-840.
[20] Johan Jacob Jijo. (2025). Modelling and Forecasting the GDP of G7 Countries Using Arima Model. Asian Journal of Economics, Business and Accounting,25(3),295-315.
[21] Anchal Dahiya,Pooja Mittal,Yogesh Kumar Sharma,Umesh Kumar Lilhore,Sarita Simaiya,Mohd Anul Haq... & Abdullah Alenizi. (2025). Hybrid parking space prediction model: integrating ARIMA, Long short-term memory (LSTM), and backpropagation neural network (BPNN) for smart city development.. PeerJ. Computer science,11,e2645.
[22] Gu Minghui. (2023). Population Prediction of China Based on ARIMA-LSTM Combined Model. Journal of Sociology and Ethnology,5(3),
[23] Li Chuan,Fang Xinqiu,Yan Zhenguo,Huang Yuxin & Liang Minfu. (2023). Research on Gas Concentration Prediction Based on the ARIMA-LSTM Combination Model. Processes,11(1),174-174.