ISSN: 1813-4890

A combination of swarm intelligent optimization algorithms Research on the prediction of employment number by forecasting model

Zixin Wang, Feng Xu, Zhenghong He, Jianzhe Cao, Yucong Li

Anhui University of Finance and Economics, School of Statistics and Applied Mathematics, Bengbu, Anhui 233000, China

*Corresponding author:2269591781@qq.com

Abstract

Employment is closely related to social stability and the stable growth of international economy. In order to scientifically predict the number of rural employment, this paper selects the data of the number of rural employment in the country from 2016 to 2022, uses ant colony optimization algorithm and solves it with Python, Matlab and other software, and builds a time series prediction model of ant colony algorithm through analysis. Finally, the accurate and efficient prediction of the number of rural employment is proposed. Firstly, this paper introduces the significance of the research and the current situation of the domestic research on employment. Secondly, it introduces the related concepts of the forecast model to be used in the analysis of the employment problem. Then, it puts forward the application of ant colony algorithm and time series prediction in the study of urban and rural employment number; Finally, the conclusion is drawn.

Keywords

Employment number; Python; Matlab; Time series prediction; Ant colony algorithm.

1. Introduction

1.1. Research Significance

Employment is fundamental to people's livelihood and is closely related to social stability and steady growth of the international economy. It bears on the vital interests of the people, on the overall situation of reform, development and stability, and on the realization of the grand goal of building a moderately prosperous society in an all-round way. Therefore, people pay great attention to the issue of urban and rural employment. Many experts have carried out research on it, but most of them only carry out theoretical research and few use statistical analysis methods. At present, the accuracy of the model of the employment number forecasting method cannot be effectively improved, resulting in unstable output results, and there are still certain limitations in use. Its initial network parameter Settings are more random, and the network output results are less stable.

The traditional deterministic prediction usually can't provide any information with uncertain factors, and accordingly, it can't measure the risk of the time series prediction object comprehensively, thus its use value and application scope are limited to a certain extent. However, the combinatorial prediction model of swarm intelligence optimization algorithm can obtain the result of time series prediction with stronger generalization performance, more accurate and stable. Therefore, the accurate combinatorial prediction of swarm intelligence optimization algorithm is very important for people to make the right decision.

The time series forecasting model based on swarm intelligence optimization algorithm can accurately and stably predict the information about the development of employment issues according to the existing data, and provide effective information support for the management departments and policy making departments in related fields to formulate more effective policies. At the same time, for a single forecasting method, the model can solve the problem of local or global optimization in the process of forecasting, effectively improve the accuracy of the model, so as to obtain more accurate and stable forecasting results.

1.2. Research status

At present, there are some deficiencies in the research on rural employment in that although the structural analysis of employment is sufficient, the research on its supply structure, as one of the strategic theories of employment development in China, has not used much space, and the research content is not sufficient, which needs to be supplemented.

With the continuous progress of science and technology, a large number of scholars have improved the effect of time series forecasting based on swarm intelligence optimization algorithm and artificial intelligence. Swarm intelligence optimization algorithm, as a branch of intelligent computing methods, has been widely concerned by the industry in recent years. The proposal of swarm intelligence optimization algorithm was initially inspired by the crowd behavior in nature. Many intelligent optimization algorithms have been invented to solve the above complex optimization problems, such as the ant colony algorithm that simulates the collective track finding of ants. Time series prediction model has been widely used in energy, finance, environment, atmosphere and other fields, and has a relatively mature theoretical and practical research foundation.

When forecasting, it tries to filter out irregular changes, highlight trend and cyclical changes, and can synthesize data features related to employment to further improve the results of time series forecasting the accuracy and reliability of the forecast.

1.3. Literature Review

At present, multiple headwinds are adding up to lower employment expectations.(State Information Center, 2022) Therefore,there are many problems in China's employment that need to be solved.

Due to the combination of factors such as the economic slowdown and the epidemic, the number of jobs available in the job market has been greatly reduced. On the other hand, graduates have high employment expectations, and some college graduates' actual ability cannot meet the needs of employers, leading to the coexistence of "recruitment difficulties" and "employment difficulties". There is also a "35 years old threshold" in society and discrimination against older unmarried women without children, which affects the employment of workers. (Fu Xiao, 2023)

Time series forecasting is a science that conforms to the needs of social and economic production management by scientific analysis of historical and current information to predict and reveal the future development trend and law of things (Diebold,1998).

In time series analysis, the traditional time series model has incomparable advantages in dealing with small samples, linear and stationary data. It includes trend extrinsic model (Rhyne,1974), regression model (Calton,1886), Auto Regressive (AR) model (Yule,1927), moving average model (MovingeAverage).

MA) (Walker,1931). However, the effect is poor when dealing with large samples, nonlinear and non-stationary data, and there are problems such as large computational cost and difficult optimization (Li Wen, Deng Sheng, Duan Yan, Du Shouguo, 2020).

Although there have been many works and theories on employment and time series forecasting, and the relevant research theories are quite mature, there are still very few

ISSN: 1813-4890

researches on in-depth analysis of existing data combined with examples. This paper attempts to use the knowledge of swarm intelligence optimization algorithm to comprehensively analyze China's urban and rural employment data to make time series combination forecasting. This project mainly applies the combined forecasting model based on swarm intelligence optimization algorithm to the research of the number of employment , and applies the above model to the forecast of the number of employment time series, in order to obtain more accurate, stable and generalized time series forecasting results.

2. Relevant knowledge reserve and data reserve

2.1. Research relevant theoretical knowledge

2.1.1. Ant colony optimization algorithm

Ant colony optimization algorithm is a population intelligence algorithm, which is a kind of intelligent behavior shown by a group of non-intelligent or slightly intelligent individuals due to interaction. The basic principle of ant colony optimization algorithm is derived from the shortest path principle of foraging in nature. According to expert observations, ants can find a shortest path from the food source to the nest without any prompting, and can spontaneously find a new shortest path after the environment changes, such as the appearance of an obstacle in that path.

The steps of the ant colony optimization algorithm areas follows:

Initialization.

Select the next node for each ant.

Random ratio rules in AS; For each ant k, the path memory vector R^ K. records the sequence numbers of all the cities that k has passed through in order of access. Let the current city of ant kbe i, then the probability that ant K will choose city j as the next object to visit is:

$$p_k(i,j) = q\begin{cases} \frac{[\tau(i,j)]^\alpha \gamma(i,j)^\beta}{\sum_{\mu\epsilon J_k(i)}[\tau(i,j)]^\alpha \gamma(i,j)^\beta}, j\epsilon J_k(i) \\ 0 \end{cases} \tag{1}$$

Update the pheromone matrix.

$$\tau(i,j) = (1-\rho).\tau(i,j) + \textstyle\sum_{k=1}^{m}\Delta\tau_k(i,j) \tag{2}$$

$$\Delta\tau_k(i,j) = \begin{cases} (C_k)^{-1}, (i,j)\epsilon R^k \\ 0, \text{ohters} \end{cases} \tag{3}$$

$$\gamma\ (i,j) = \frac{1}{d_{ij}} \tag{4}$$

Here mis the number of ants, ρ is the evaporation rate of the pheromone, specified $0\le \rho\le 1$, in AS is usually set to ρ =0.5, Δτij is the amount of pheromone released by the kant on the side it passes, it is equal to the reciprocal of the length of the ant k epicycles build path. Ck represents the pathlength, which is the sum of the lengths of all sides in Rk.

Check the termination condition

If the maximum algebra MAX_GEN is reached, the algorithm terminates and goes to step (5); Otherwise, reinitialize all the ant Delt matrix with all elements initialized to 0, the Tabu table emptied, and all the city nodes added to the Allowed table. Select their starting locations at random (you can also specify them manually). Add the starting node to Tabu, remove the starting node from Allowed, and repeat steps (2), (3), (4).

Output the optimal value

ISSN: 1813-4890

2.1.2. Time series prediction

A time series is a sequence of data that is chronological, changes over time, and is related to each other. Here are the steps:

For each moment of time to be predicted, where is the number of previous observations,N is the maximum possible length of the part of the time series that the sequence takes up.x_t, tÍ$(t_0; t_0 + N)$

From the set of typical sequences, find all the known predicted parts of the corresponding time series and the sequences containing each element. xt

If, where L is the length of the portion of the sequence in the time series, x_{t+i}^{old}andx_{t+i}^{new}, respectively, represents the initial number of times that the previous sequence and the current sequence enter the time series$\forall i: 0 < i < L: \left|x_{t+i}^{old} - x_{t+i}^{new}\right| < D_{max}$

For the prediction problem, the sequence is collected from the set generated by (2). xt

Select the group that minimizes the standard deviation of the group sequence values and the known predicted portion of the time series.

Calculate the predicted value as the average of the group series. xt

2.2. Research Data

In order to predict the number of rural employment, the data I have listed below are all from the number of rural employment (10,000) from 2016 to 2022 in the Statistical Data of China Economic Network.

Table 1 Number of rural employment in 2016-2022

Year	Number of employed people in rural areas(10,000)
2016	34194
2017	32850
2018	31490
2019	30198
2020	28793
2021	27879
2022	27420

3. Application of swarm intelligence optimization algorithm in time series prediction of rural employment number

3.1. Application Examples

This paper uses the data of China Economic Network from 2016 to 20227 to establish a swarm intelligent optimization algorithm to predict the number of rural employment in the coming years. Table 1 shows the specific number of rural employment.

Table 2 The number of rural employment

Year	2016	2017	2018	2019	2020	2021	2022
Employment numbers	34194	32850	31490	30198	28793	27879	27420

Based on the data in the table, it is estimated that the number of rural employment will decline gradually in the future. In order to better predict the future rural employment situation, the formula generated by accumulation is used to transform the series into a non-negative and increasing series.

$$x^{(1)}(k)=\sum_{i=1}^{k}x^{(0)}(i) \tag{5}$$

Where $x^{(1)}(k)$ is the new sequence generated after summation, and $x^{(0)}(i)$ is the initial sequence. The data in Table 1 are obtained by the method of accumulation generation, as shownin Table 3.

Table 3 Generates a new series of numbers

Serial Number	1	2	3	4	5	6	7
X(0)(i)	34194.3	32850.4	31490.2	30198.3	28793.2	27879.4	27420.6
X(1)(i)	34194.3	67044.7	98535.0	128733.3	157526.6	185406.0	212826.7

Construct a new sequence of first-order differential equations with the year: $\frac{dx^{(1)}(t)}{dt}+ax^{(1)}(k)=\mu$. For easier operation, the first order differential equation is converted to $x^{(0)}(t)=-ax^{(1)}(t)+\mu$. Since the parameters a and are unknown, the least square method and the minimization error are used for parameter estimation. Considering that the first order differential equation contains $\frac{dx^{(1)}(t)}{dt}$, the new sequence $x^{(1)}(k)$ is modified to mean generating sequence $z^{(1)}(t)$:

$$z^{(1)}(t)=0.5x^{(1)}(t)+0.5x^{(1)}(t-1) \tag{6}$$

The matrix form contained in the least squares estimation equation is as follows:

$$B=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}=\begin{bmatrix} -\frac{1}{2}\left[x^{(1)}(2)+x^{(1)}(1)\right] & 1 \\ -\frac{1}{2}\left[x^{(1)}(3)+x^{(1)}(2)\right] & 1 \\ \vdots & 1 \\ -\frac{1}{2}\left[x^{(1)}(n)+x^{(1)}(n-1)\right] & 1 \end{bmatrix}\begin{bmatrix} a \\ \mu \end{bmatrix} \tag{7}$$

Use the least squares parameter column to satisfy, where. $P=(\hat{a},\mu)^T=(BB^T)^{-1}B^TY$,which $Y=\begin{bmatrix} x^{(0)}(2) \\ x^{(0)}(3) \\ \vdots \\ x^{(0)}(n) \end{bmatrix}$ .The estimated values of a are 0.0381 and 34597.6722 respectively by substituting the data to solve a, μ.

Therefore, the prediction model is $\overline{x}^{(1)}(i+1)=-873881e^{-0.0381i}+908075.4$.

3.2. Residual Analysis

If the actual value (i) is $x^{(0)}(i)$,and the simulated value (i) is $\overline{x}^{(0)}(i)$, then the residual is $\varepsilon^{(0)}(i)=x^{(0)}(i)-\overline{x}^{(0)}$. As can be seen from the table, there are only small deviations (less than 2% standard errors) between the simulated and actual values for 2016, 2017, 2018, 2020, 2021, and 2022, with the relative error reaching 2.76% only for 2019. The relative average error over these seven years is. The test results show that the prediction model has a reasonable structure and high accuracy.

Table 4 residual analysis results

Serial number（i）	Year	Theactual value$x^{(0)}(i)$	The simulated value$\overline{x}^{(0)}(i)$	the error$\varepsilon^{(0)}(i)$	Relative error $\psi^{(0)}(i)/\%$
1	2016	34194.3	34194.3	0	0
2	2017	32850.4	32337.9	512.47	-1.56
3	2018	31490.2	31877.5	387.33	1.23
4	2019	30198.3	31031.8	833.47	2.76
5	2020	28793.2	29181.9	388.71	1.35
6	2021	27879.4	28180.5	301.10	1.08
7	2022	27420.6	27025.7	394.86	-1.44

3.3. Correlation Analysis

Based on the actual value (i)$x^{(0)}(i)$, the proposed value (i)$\overline{x}^{(0)}(i)$ and the residual (i)$\varepsilon^{(0)}(i)$, the association coefficient is used to define $\xi(i)=\frac{\min(\Delta(i))+\rho\max(\Delta(i))}{\Delta(i)+\rho\max(\Delta(i))}$,for the ith data, the correlation coefficient can be calculated, and the results are shown in the Table 5.

Table 5 association coefficient data

Serial number（i）	1	2	3	4	5	6	7
$\xi(i)$	1	0.327	0.455	0.510	0.536	0.496	0.420

According to the above table, the corresponding correlation degree is obtained:

$$\xi=\frac{1}{n+1}\sum_{i=1}^{n}\xi(i)=(1+0.327+0.455+0.510+0.536+0.496+0.420)/6=0.624. \quad (8)$$

From the formula: $\xi=\frac{1}{n+1}\sum_{i=1}^{n}\xi(i)$ =0.624>0.5,so, in general,when ρ=0.5 and ξ=0.624, the results are satisfactory.

3.4. Post-residual Analysis

Post-residual analysis is to evaluate the validity of the prediction model by analyzing whether the residual sequence of the prediction model has the statistical characteristics of randomness, stationarity and independence. In the time series prediction method based on ant colony optimization, we can carry out post-residual analysis by the following methods:

Calculate the residual sequence of the model: Apply the prediction model to the historical data to obtain the corresponding predicted value. Then, the residual sequence is obtained by calculating the difference between the observed and predicted values.

Test the randomness of the residual sequence: statistical methods (such as Ljung-Box test) are used to test whether the residual sequence has significant autocorrelation. If there is no significant autocorrelation in the residual series, it indicates that the prediction model has successfully captured unexplained random fluctuations in the time series data.

Test the stationarity of the residual series: By testing the stationarity of the residual series, it is possible to assess whether the prediction model can accurately capture the overall trend and fluctuation of the time series data. Common methods include unit root tests (such as ADF tests) and stationarity graph analysis.

Testing the independence of the residual sequence: Testing whether the residual sequence is independent is one of the important indicators to evaluate the prediction model. Common methods include Durbin-Watson statistics and Ljung-Box tests. If there is no significant autocorrelation and heteroscedasticity in the residual series, it indicates that the prediction model can fully explain the fluctuations in the time series data and has good prediction ability.

The method of calculating the residual sequence of the model is used. According to the basic principle of grey system, the variance ratio $C = S/S_0 = 0.1049$ can be obtained by calculating the mean square error S_0 of the original time series $x^{(0)}$ and the mean square error S of the residual series $\varepsilon^{(0)}$.

Because the minimum error probability $p = \{|\varepsilon^{(0)} - \bar{\varepsilon}^{(0)}|\}$, it can be seen from the division of model accuracy in Table 1 that the prediction model has good accuracy and is suitable for the prediction of rural employment.

The results of the above tests show that the grey system model constructed in this paper is qualified and has high accuracy, which is suitable for the prediction of rural employment number. Based on the forecast model, we can see that the number of rural employed people shows a decreasing trend in the seven years from 2016 to 2022, and the annual reduction rate is about 4%, as shown in Table 5.

Table 6 Number of rural employment from 2016 to 2022 (10,000)

A given year	Number of people	A given year	Number of people	A given year	Number of people	A given year	Number of people
2016	34194	2018	31490	2020	28793	2022	27420
2017	32850	2019	30198	2021	27879		

The main reasons for the continuous reduction of rural employment are:

The rural employment industry is single and the position is insufficient. In the development of urban integration, infrastructure development is in the first place, most of the infrastructure projects are also in accordance with the design and planning, and can not provide multi-industry, meet the needs of multiple jobs in one step, and the current new infrastructure investment is more market and government to promote, such as the urbanization of the southern Henan region is basically to project investment, government support back to rural entrepreneurship. The positions that can be provided, such as service posts, factory posts, infrastructure posts, core technical posts, etc., are still mainly external employment.

The average salary in rural areas is not high. Regardless of village infrastructure construction or talent introduction salary standard is still implemented according to local conditions, there are different salary standards according to different posts, but the salary standard of each post can not be compared with that of the city, which has a certain relationship with the scale and nature of township enterprises, so it has caused a large reduction in the number of employment. There is a big gap in the public service system in rural areas.

The large gap between rural public service systems is first reflected in the imbalance of educational resources. At present, high-quality and efficient educational resources are mostly concentrated in cities and counties, and rural primary and secondary schools have become the short board of education. Second. The distribution system of medical resources is also unbalanced, and high-quality medical resources are concentrated in first-tier, second-tier, and third-tier cities in an inverted pyramid, making rural medical services far behind urban counties. Multiple reasons promote the continuous influx of rural talents from the countryside to the city, and urban talents are reluctant to return to the countryside, the lack of career development prospects in the countryside, due to the lack of resources caused by the vicious circle will lead to the loss of personnel and slow development.

4. Conclusion

Starting from the problem of the number of national employment , this project proposes a prediction model based on swarm intelligence optimization algorithm. Ant colony algorithm is used to extract the relevant information of the singular attractor topology from a given time series, and the extracted information is used to predict the number of national rural employment. The experimental results show the feasibility and usability of this method.

Based on this model, the forward propagation and error reverse propagation of the national employment data in the past 30 years are used to update the parameters and minimize the loss. This provides an effective basis for the government to formulate the strategy of rural revitalization and development; And for investors to make scientific and rational investment and management decisions when considering rural development.

Acknowledgements

Anhui University of Finance and Economics Undergraduate Research Innovation Fund Project Support (XSKY23237).

Key project of scientific research plan of colleges and universities in Anhui Province 2022AH050565.

References

[1] Wang Su. Research and Application of Time series Prediction Algorithm based on Deep Learning [D]. University of Electronic Science and Technology of China,2023.

[2] Tian Shaolong. Application of principal component analysis and time series forecasting in the study of employment numbers in China [D]. Soochow University,2016.

[3] Su Dawei. Research on Urban-Rural Integrated Employment Based on Grey System Theory [D]. Xi 'an University of Technology,2007.

[4] Zhou Q R. Opportunity and Ability: Employment and Mobility of Rural labor Force in China [J]. Managing the World,1997.

[5] Yu Enhai, Liu Dezhong. Review of research on Employment in China [J]. Economic Trends,1999.

[6] Ni Songxin. The Current Situation and development trend of Urban and rural Labor employment in China [J]. China Labor Science,1986.

[7] Hu Angang. Analysis of Employment Situation in China [J]. Management World,1997.

[8] Wang Shaoguo. Analysis on the impact of China's economic growth and industrial structure upgrading on urban employment [J]. Current Finance,2005.

[9] Yu Jinhua, Yang Baowei. Multivariate Statistical Analysis and Application [M]. Guangzhou: Sun Yat-sen University Press,2006.

[10] Tsay,Ruey S.Analysis of Financial Time Series[M].2005.

[11] Swami,A., and J.M.Mendel.ARMA parameter estimation using only output cumulants[J].IEEE Trans.ASSP,1990,38:278-305.

[12] George E.P.Box.Time Series Analysis:Forecasting and Control[M].2009

[13] I.T.Jolliffe.Principal Component Analysis[M].2002.

[14] Gavin W Jones and Pravin Visaria.Urbanization in large developing countries:China, Indonesia, Brazil and India, New York:Clarendon Press.

[15] Ünler, A. Improvement of Energy Demand Forecasts Using Swarm In-telligence: the Case of Turkey with Projections to 2025[J]. Energy Poli-cy, 2008 ,(36).

[16] Wang, J., Chi, D., Wu, J., Lu, H. Chaotic Time Series Method Com-bined with Particle Swarm Optimization and Trend Adjustment forElectricity Demand Forecasting[J].Expert Systems with Applications,2011,(38).

[17] Weng, S. S., Liu, Y. H. Mining Time Series Data for Segmentation byUsing Ant Colony Optimization[J]. European Journal of OperationalResearch, 2006,(173).

[18] Wong, W. K., Xia, M., Chu, W. C. Adaptive Neural Network Modelfor Time-series Forecasting[J]. European Journal of Operational Re-search, 2010,(207).

[19] Zhao, L., Yang, Y. PSO-based Single Multiplicative Neuron Modelfor Time Series Prediction[J].Expert Systems with Applications,2009, (36).