Research on prefecture-level city's birth rate prediction based on AE-LSTM network

Mingfu Xue [1, a], Pei Huang*, [2, b]

[1] Johns Hopkins University, USA;

[2] Bank of Changsha, China.

[a] xue736248@163.com, [b] huangpeisgxty@163.com

Abstract. In recent years, China's birth rate has continued to decline, leading to a series of population and economic problems. Although the country has gradually liberalized its Family Planning Policy and started implementing a Three-Child Policy with supporting fertility measures, it has not significantly increased the birth rate. To further implement the policy of encouraging childbirth at the prefecture-level city, this article proposes an AE-LSTM model for predicting the birth rate of the prefecture-level city, combining multidimensional time-series data that affect the birth rate of the population, which achieves accurate prediction of the birth rate of the prefecture-level city, providing a certain decision reference for prefecture-level city governments to formulate more accurate and reasonable childbirth encouragement policies. At the same time, experiment results show that the method proposed in this article has higher accuracy and better generalization performance compared with other mainstream methods.

Keywords: Birth Rate Prediction; Economic Development; AE-LSTM.

1. Introduction

In the year of universal "Universal Two-Child" was officially implemented on January 1, 2016, 17.86 million people were born in China, with a birth rate of 12.28‰. However, since 2017, China's birth rate has been declining year by year, and the number of births has plummeted. According to the seventh national census report in 2020, China's newly born population was 12 million, and the birth rate was only 8.5‰, hit a record low. Moreover, predictions show that China's population size will shrink sharply from 2050 onwards, which lead to great challenge on labor supply. And the disappearance of demographic bonus will make China lose its comparative advantage in the global economy market and weaken its comprehensive national strength[1]. In May 2021, the Chinese government adopted the "Decision on Optimizing the Birth Policy to Promote Long-term Sustainable Population Development", further implementing the three-child policy.

The continued decline in the birth rate will further accelerate China's population aging and lead to great challenge on labor supply. Combined with the current trend of declining birth rates, the scale of China's labor force will continue to decrease, and a series of social problems that come with it will become obstacles to national and social development, such as population crisis[2], labor shortage[3], declining productivity, heavier household burdens[4], GDP growth rate decline[5], and reduced innovation effects[6]. As the overall policy documents at the national level are unable to quickly solve the problem of birth rates decline each city should comprehensively promote encouraging fertility policies based on their own different characteristics. Therefore, predicting and analyzing the birth rate especially at the prefecture level has very important theoretical value and practical significance.

Currently, most birth rate predictions are based on dynamic regression models (ARIMAX), which mainly use stationary output variables and multiple input variables to construct models. H et al.[7] used US birth rate data from 1975 to 2001 to accurately predict US birth rates from 2002 to 2008 using ARIMA models. Hasan et al.[8] constructed a multivariate stepwise linear regression model to predict Bangladesh's birth rate. However, such methods rely too much on manual prior knowledge and cannot extract deep clues and coupling relationships between variables. With the rapid development of machine learning, research on population birth rate prediction based on machine learning and neural network technology has gradually become mainstream and a hot topic. Jia et al.[9] and Wu et al.[10] used BP neural networks to predict birth rates. However, traditional machine learning methods cannot accurately extract coupling features between multiple variables and have too many parameters. Therefore, this article proposes an AE-LSTM network to extract mixed features of eleven factors with the highest correlation with birth rates in prefecture-level cities and achieve high-precision birth rate prediction.

2. Method

2.1 overview

Circular neural networks (RNNs) are a type of machine learning algorithm used to process time series data. Hochreiter et al.[11] improved the general RNN by proposing the long short-term memory (LSTM) model. The basic unit of the LSTM model is the memory module, which includes four neural network layers: the cell state, forget gate, input gate, and output gate, as opposed to the simple tanh layer repeated in general RNNs. The LSTM model solves the problem of long-term dependency in general RNNs. It can handle and predict important events with long time intervals and delays in event sequence and applies distant contextual information to the current moment. Therefore, using LSTM as a core model for predicting prefecture-level city's birth rate can fully connect information data from different years in the same region, establish cross-scale feature connections, and achieve better prediction results.

However, traditional LSTM models can only handle single-variable time series data and often fail to effectively extract deep coupling clues and features between variables when dealing with multi-variable time series data. Therefore, this article introduces an autoencoder structure before the LSTM to encode high-dimensional data into low-dimensional data. The autoencoder structure can automatically extract and learn coupling features and deep clues from original multi-variable data and has found extensive applications in data compression and feature extraction projects

2.2 the proposed network

The model used in our article is presented in Figure 2.1. After conducting a thorough literature review[12][13][14][15], we have selected eleven factors that have the highest correlation with the prefecture-level city's birth rate, such as housing prices, the percentage of eligible women, and prefecture-level city's GDP levels, as the multidimensional input variables for the network model. These variables are processed through an auto-encoder module, a LSTM module, and two fully connected layers to ultimately output the predicted values for the prefecture-level city's birth rate.

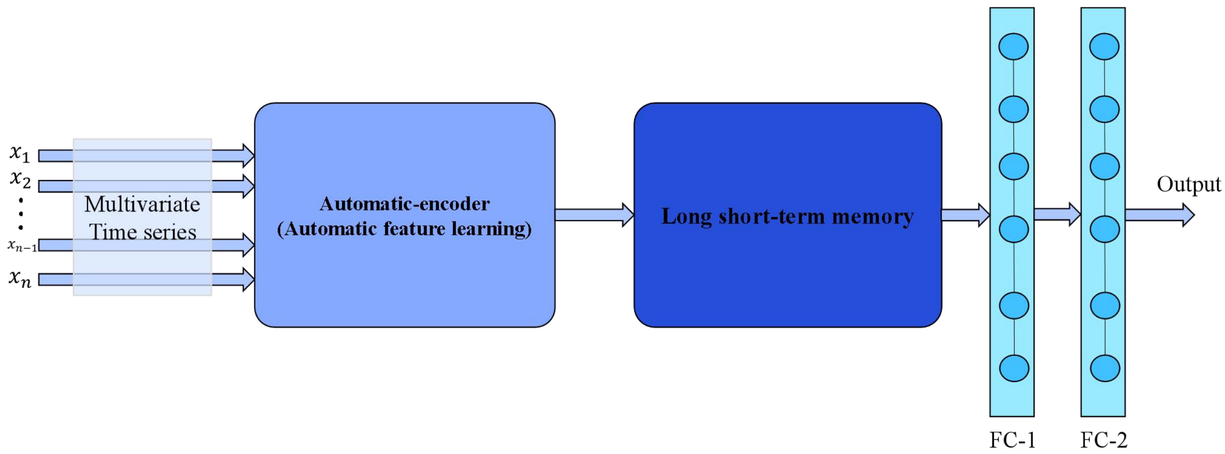

Fig. 2.1 The structure of the network proposed in this article.

The auto-encoder[16] architecture used in this article is shown in Figure 2.2, consisting of input, hidden, and output layers connected via fully connected layers. The input and output layers have the same scale (number of network nodes), while the scale (number of network nodes) of the hidden layer is smaller than that of the input and output layers. To achieve a high degree of nonlinear structure size reduction for multidimensional input data, this framework requires a multi-layer encoding and decoding structure (set to 8 in this article), with each layer structure outputting as follows:

$$y_i^{(1)} = \sigma(W^{(1)}x_i + b^{(1)}) \tag{1}$$

$$y_i^{(k)} = \sigma(W^{(k)}y_i^{(k-1)} + b^{(k)}), k = 2, \ldots, K \tag{2}$$

Where (1) and (2) are the outputs of the first and k-th hidden layers, respectively. σ represents the sigmoid activation function. $W^{(k)}$ and $b^{(k)}$ represent the weight matrix and the bias vector of layer k. K refers to the number of hidden layers, and $y_i^{(k)}$ denotes the feature maps extracted by the hidden layers.

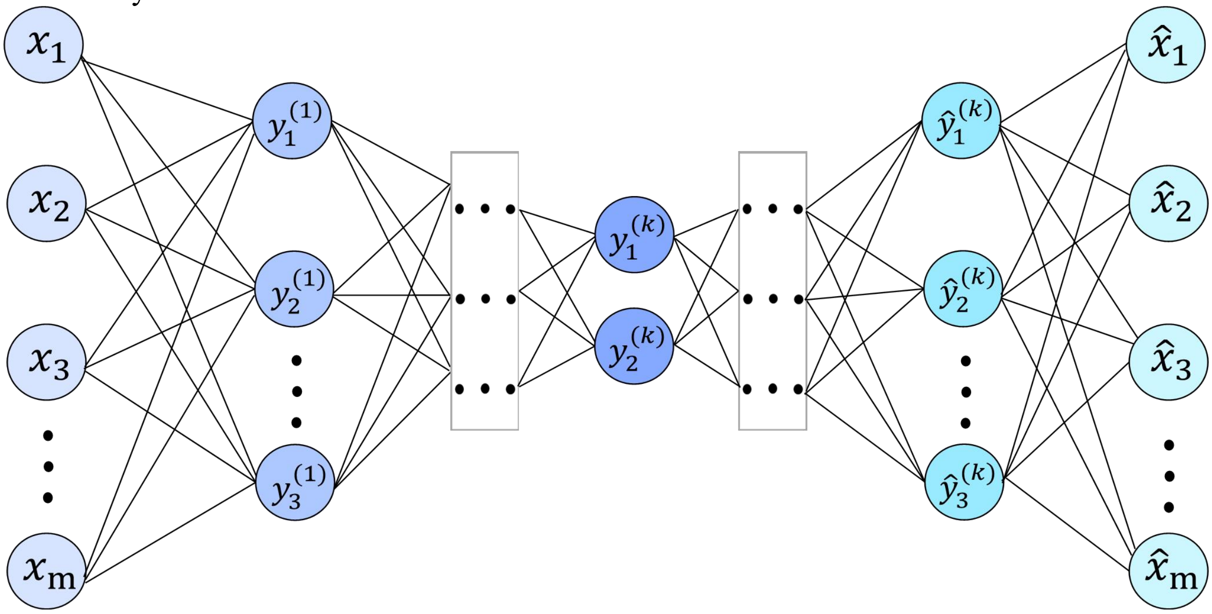

Fig. 2.2 The structure of the Auto-encoder.

The LSTM architecture used in this article is shown in Figure 2.3. An LSTM unit consists of three gate units. Long short-term memory (LSTM) is a well-known deep learning architecture used for processing time series data. It incorporates gate mechanisms and internal states to enable selective accumulation of historical information and retention of new information[17]. Figure 2.3 illustrates the structure of an LSTM unit. The LSTM unit comprises three gates, namely the forget gate f_t, input gate i_t, and output gate o_t. The forget gate f_t decides how much historical information to retain in the internal state when new information is introduced. The input gate i_t determines which new information should be updated to the internal state, while the output gate o_t selects which updated information can be passed as input to the network for the next time step[18]. When dealing with time series prediction problems, the LSTM network has faster convergence speed and higher accuracy compared to other neural network models.

2.3 Loss function and evaluation index

To further optimize the model parameters, it is necessary to minimize the reconstruction error. We have used mean squared error as the loss function, which is expressed as follows:

$$L(x, z) = \frac{1}{n}\sum_{i=1}^{n} (x_i - \hat{x}_i)^2 \tag{3}$$

Where $\hat{x}_i$ represents the predicted value of the prefecture-level city's birth rate, and x_i represents the true value.

In this article, we have evaluated the predictive performance of the model using two evaluation metrics: root mean squared error (RMSE) and coefficient of determination (R-squared score).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (|y_i - \hat{y}_i|)^2} \tag{4}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{5}$$

Where $\hat{y}_i$ represents the predicted value, y_i represents the real value, and n represents the number of data points. A smaller RMSE value indicates a smaller prediction error of the prefecture-level city's birth rate prediction model. The closer the R-squared value is to 1, the better the fit between the predicted prefecture-level city's birth rate and the actual value.

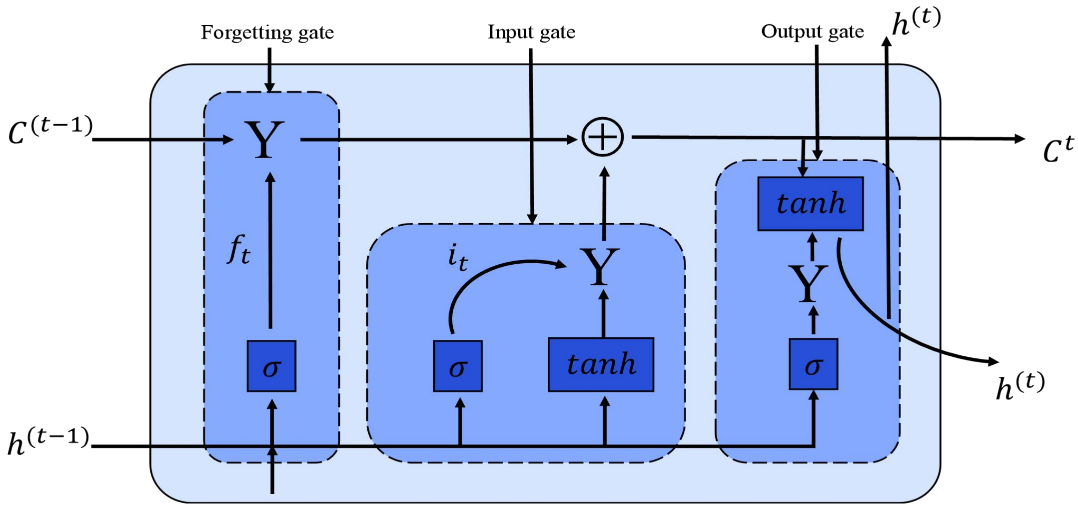

Fig. 2.3 The structure of LSTM.

3. Experiment and evaluation

3.1 Dataset

This article collected comprehensive data from 361 prefecture-level cities in China over four years (2005, 2010, 2015, and 2019) published by the National Bureau of Statistics as experimental evaluation data to verify the effectiveness of the AE-LSTM-based spatiotemporal correlation prediction method. We divided the data into training set, validation set, and test set in an 8:1:1 ratio.

3.2 Quantitative experiments

Based on industry-related literature analysis, we selected eleven factors that have the greatest impact on the prefecture-level city's birth rate as inputs to the network model, which are GDP, PBR(index of the impact of policies on childbirth), registered population relative to de jure population, primary school teachers per 10,000 people, number of practitioners per 10,000 people, number of university students per 10,000 people, old-age dependency ratio, Per capita social security and employment expenditure, housing price relative to GDP, female education level, and percentage of employed women.

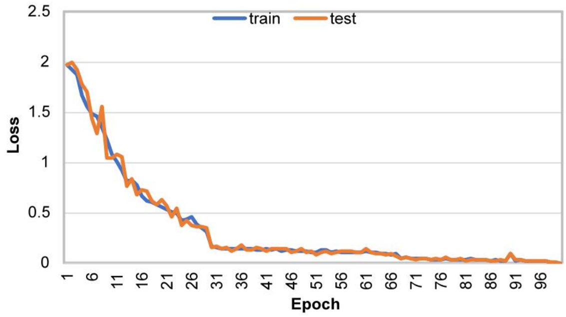

Fig. 3.1 The change in the model loss of the training set and test set.

Adam[19] is used in this article as the optimization algorithm and the network is implemented in the Pytorch[20] framework. Figure 3.1 shows the change in the model loss of the training set and test set with the increase of training iterations during the training process. It can be seen that the decrease in the loss for both the training set and test set is almost the same and close to 0. The loss of the training set is close to 0, indicating that the auto-encoder has outstanding performance in learning the data features of prefecture-level city's birth rate in China. The loss of the test set is also close to 0, indicating that the model has strong robustness and generalization ability, which can predict future birth rates with high accuracy.

In addition, this article also compared the AE-LSTM model with other popular algorithms used in prefecture-level city birth rate prediction tasks and other multi-dimensional time series data prediction tasks, such as ARIMAX and CAE. Figure 3.1 shows the prediction results of the above methods and Table 1 shows the comparison results of root mean squared error (RMSE) and coefficient of determination (R-squared score). The AE-LSTM model used in this article has the smallest RMSE value, indicating that our model has the smallest prediction error for prefecture-level city's birth rate. At the same time, the R-squared value of our method is closest to 1, indicating that our predicted prefecture-level city's birth rate has the best fit with the actual value.

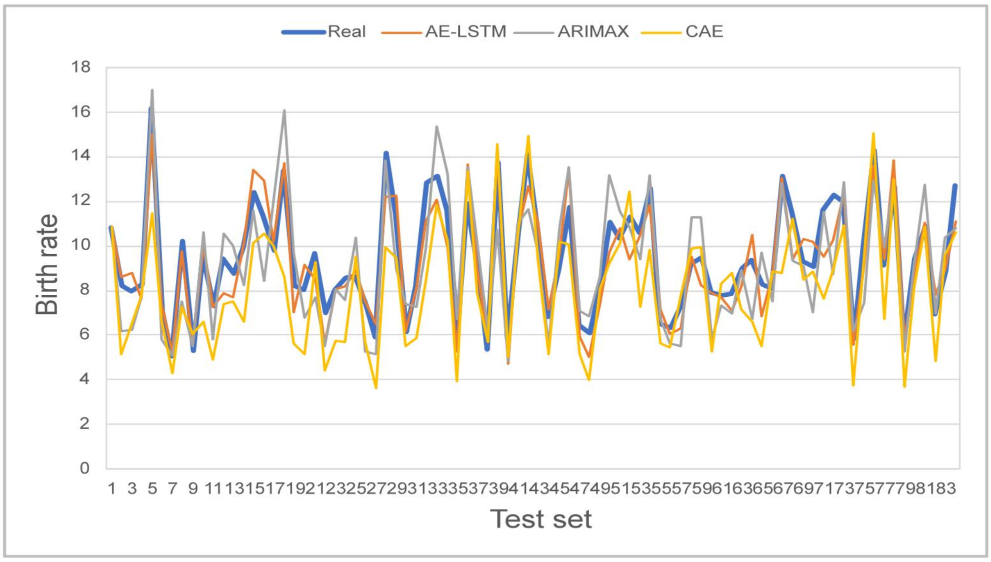

Fig. 3.2 Comparison Results of Three Methods for Birth Rate Prediction with Real Values.

Table 1. Comparing the Evaluation Metrics of Three Models for Birth Rate Prediction

Method	RMSE	R^2
ARIMAX	0.0027	1.9349
CAE	0.0012	1.8827
AE-LSTM(ours)	0.0006	1.1013

4. Conclusion

Based on the multi-dimensional time series structure of China's prefecture-level city's birth rate data, this article established an AE-LSTM model for accurate prediction of China's prefecture-level city's birth rate. Experiment results show that the AE-LSTM model used in this article has higher fitting and prediction accuracy for China's prefecture-level city's birth rate than the traditional model, and has stronger robustness and generalization performance. In the context of a continuous decline in the birth rate, from the perspective of national strategy, this research can provide a certain decision-making reference for prefecture-level city governments to formulate more accurate and reasonable birth encouragement policies.

In future research, we can further improve the prefecture-level city's birth rate prediction model. On the one hand, we can try to transfer our model to foreign city for prediction of birth rate to analyze the impact of different national conditions and policies on prefecture-level city's birth rates. On the other hand, we can continuously improve the model by using more advanced networks and modules to further improve the accuracy and generalization ability of the model.

References

[1] Z. Ren. China Population Forecast Report 2021 Edition(in Chinese)

[2] Y. Jin. The Low Fertility Trap: Theory, Evidence, and Implications. Population Studies(in Chinese).2014,38(01)：3-17.

[3] F. Cai. Population transition, demographic dividend, and Lewis turning point. Economic research(in Chinese).2010,45(04)：4-13.

[4] Alexander W. Butler and Hanyi Yi. Aging and public financing costs： Evidence from U.S. municipal bond markets. Journal of Public Economics. July 2022.

[5] Maestas, Nicole, Kathleen J Mullen, and David Powell. American Working Conditions Survey： Focus on Older versus Younger Workers.2016.NBER working paper.

[6] J. Zheng. D. Zhu. Z. Guan. The impact of the labor force and aging on economic growth: An empirical study based on nine provinces and cities in China. China's soft science(in Chinese).2014(04)︰149-159.

[7] H Baghestani, M Malcolm. Factors predicting the US birth rate. Journal of Economic Studies,2016,43(3): 432-446.

[8] HASAN, Md Mahmudul, et al. Investigation of influential factors towards predicting birth rate in Bangladesh. In: 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE). IEEE, 2020. p. 1-6.

[9] N. Jia. H.Hu. Y. Bai. Population forecasting based on BP neural network. Journal of Shandong University of Technology: Natural Science Edition(in Chinese). 2011, 25(3):3.

[10] J. Wu. Prediction of Birth rate in China under three-child Policy based on neural network. 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 1652-1655.

[11] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," in Neural Computation, vol. 9, no. 8, pp. 1735-1780, 15 Nov. 1997

[12] Murat, F, Iyigun. Timing of childbearing and economic growth[J]. Journal of Development Economics, 2000.

[13] Tan P L. The impact of school entry laws on female education and teenage fertility[J]. Journal of Population Economics, 2016, 30(2): 1-34.

[14] Feijten P, Mulder CH. The timing of household events and housing events in the Netherlands︰ a longitudinal perspective. House Study,2002,17(2):773–792.

[15] Barmby T and A Cigno. A sequential probability model of fertility patterns. Journal of Population Economics,1990:31-51.

[16] A. Chevrot, A. Vernotte and B. Legeard, "CAE: Contextual auto-encoder for multivariate time-series anomaly detection in air transportation.", Computers & Security, vol. 116, pp. 102652, 2022.

[17] X. Qiu, "Neural Networks and Deep Learning", Chinese Journal of Informatics, 2020.

[18] Y. Wu and X. Ma, "A hybrid LSTM-KLD approach to condition monitoring of operational wind turbines.", Renewable Energy, vol. 181, pp. 554-566, 2022.

[19] KINGMA, Diederik P.; BA, Jimmy. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[20] PASZKE, Adam, et al. Pytorch: An imperative style, high-performance deep learning library. Advances in neural information processing systems, 2019, 32.