

Research on Librarian Demand Prediction Based on the GM (1, 1) Model and BP Neural Network Combined Model



Leilei Peng, Ying Liu, and Ke Chen

Abstract With the development of science and technology, the automation level of library is constantly improving, and the number of librarians is decreasing year by year. As one of the most important human resources to promote the sustainable development of library, librarians are particularly important, therefore, it is necessary to research its changing trend. This study constructs a GM (1, 1)-BP Neural Network combined model, and takes the library of Sichuan University as case study. The simulation results show that, compared with a single prediction model, the GM (1, 1)-BP Neural Network combined model has higher prediction accuracy and smaller errors, which can further improve the prediction accuracy.

1 Introduction

Human resources are the most important and valuable resources in promoting the sustainable development of libraries [1], and librarian is the core component of human resources. Collections, equipment, physical space, virtual space, etc. are vital to the development of the library, but it is the librarians that ultimately play a decisive role. In recent years, with the development of information technology, traditional librarians continue to learn new technologies and acquire new capabilities. Some traditional librarians are transformed into subject librarians, they integrate into the research group and teaching team to carry out subject analysis, content demonstration, achievement evaluation and so on [2]. Some traditional librarians are transformed into digital librarians to carry out data access service, data management and so on [3].

The development of information technology has brought tremendous changes to the librarian. Taking China as an example, with the continuous expansion of enrollment in universities, the pressure of various work in the library is increasing, while the number of librarians is decreasing instead of increasing. According to the

L. Peng · Y. Liu · K. Chen (✉)

Library, Sichuan University, No. 24 South Section 1, Yihuan Road, Chengdu 610065, China

e-mail: chenkechenke@scu.edu.cn

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2023

115

L. C. Jain et al. (eds.), *Advances in Wireless Communications and Applications*,

Smart Innovation, Systems and Technologies 299,

https://doi.org/10.1007/978-981-19-2255-8_12

data released by the Steering Committee for Academic Libraries of China, in 2013, the number of librarians in 484 universities in China was 21,685, which reduced to 19,045 in 2019 [4]. Therefore, it is necessary to predict the changing trend of the librarian, so as to scientifically and reasonably allocate and regulate this key resource, and promote the development of university libraries under the background of the information era.

Based on the construction of GM (1, 1) model, BP neural network model and the GM (1, 1)-BP neural network combined model, this study conducted a case study of Sichuan University Library, and fitted the number of librarians from 2010 to 2020 and verified the combined model, and find out the law of librarian change, so as to optimize the allocation of human resources.

2 Research Method

GM (1, 1)-BP neural network combined model is constructed for research. GM (1, 1) model and BP neural network have their own advantages and disadvantages, GM (1, 1) model requires a small amount of data, and BP neural network is suitable for dealing with objects with complex internal structure. The use of one model only for the prediction may lead to deviations between the predicted value and the real data. Therefore, this study combines the two models for comprehensive prediction.

Based on the prediction of GM (1, 1) model and BP neural network, this study obtains the final weight by determining the weight of the single prediction model. According to the scholar's previous research [5], residual sum of squares (RSS) can be used to reflect the accuracy of the model. The smaller the RSS, the higher the accuracy of the model. Therefore, we determine the weight of the combined model by minimizing the RSS.

For w_i is the weight of the i th prediction, $i = 1, 2, \dots, j$. In this study, we use two prediction models, the first is GM (1, 1) model, and the second is BP neural network. The error corresponding to the combined sequence at time t is δ_t , the corresponding error of the i th prediction at time t is $\delta_{i(t)}$. Q is the RSS of combined prediction errors.

The steps of constructing combined model are:

$$\delta_t = L_{(t)} - x_{(t)} \quad (1)$$

where, $L_{(t)}$ is the predicted value, and $x_{(t)}$ is the real data.

Constrained objective function Q :

$$\text{Min } Q = \sum_{t=1}^n |\delta_t^2| = \sum_{t=1}^n \left| \sum_{i=1}^j (w_i \delta_{i(t)})^2 \right| \quad (2)$$

Subject to:

$$\sum_{i=1}^j w_i = 1, 0 \leq w_i \leq 1 \quad (3)$$

The combined model is:

$$y_t = w_1 L_1(t) + w_2 L_2(t) + \dots + w_j L_j(t) \quad (4)$$

3 Empirical Study

3.1 Brief Introduction of Research Case

This research takes the Sichuan University Library as the case study. Sichuan University Library is the library with the longest history and the largest scale of documents in Southwest China, it is composed of Arts and Science Library, Engineering Library, Medical Library and Jiang'an Library. The library has a total of 8.17 million paper documents, 312 electronic document databases, 3.68 million electronic books, 120,000 electronic journals, and 190,000 h of audio and video. In 2010, the library of Sichuan University collected 5.94 million documents, which increased to 8.17 million in 2020, and the annual expenses increased from about 17.95 million in 2010 to about 43.01 million in 2020. The substantial increase in the collection of documents and expenses will inevitably lead to an increase in the work content, but the number of librarians has dropped from 225 to 162 in the past 11 years. Therefore, Sichuan University Library should allocate human resources reasonably and explore a scientific human resources management system.

3.2 Data Source and Standardization

This study selects the data of Sichuan University Library from 2010 to 2020 as the sample data, data source is the statistical data of Steering Committee for Academic Libraries of China. The explanatory variables are the purchase cost of literature resources (x_1), the purchase cost of electronic resources (x_2), the purchase cost of paper resources (x_3), and the annual expenses (x_4), and the explained variable is the number of librarians (y). The explanatory variables are closely related to the explained variable, the higher the costs, the greater the workload of the library, the more librarians are needed. The purchase cost of literature resources includes the purchase cost of electronic resources, paper resources, multimedia materials, etc.

Table 1 Data standardization

Year	y	x ₁	x ₂	x ₃	x ₄
2010	1.0000	0.0000	0.1509	0.0000	0.0000
2011	0.9048	0.0946	0.2511	0.0195	0.0372
2012	0.8413	0.0222	0.0986	0.1828	0.0234
2013	0.6667	0.1622	0.2515	0.2429	0.1131
2014	0.5873	0.6834	0.7922	0.3357	0.7684
2015	0.4762	0.9060	0.9220	0.7527	0.8478
2016	0.3968	0.1130	0.0000	0.7970	0.1173
2017	0.3810	1.0000	0.9179	1.0000	0.8731
2018	0.1905	0.8266	1.0000	0.2498	0.7423
2019	0.1111	0.9172	0.9618	0.6786	1.0000
2020	0.0000	0.7708	0.9209	0.3032	0.7598

The annual expenses include document purchase cost, document binding and repair resource cost, equipment asset purchase cost, equipment and facility maintenance cost, office cost, etc.

The dimensions of each index are different and cannot be directly compared, so it needs to be dimensionless, and the range transformation method is used to make the index dimensionless, as shown in formula 5:

$$y_{ij} = \frac{x_{ij} - \min_{1 \leq i \leq n} \{x_{ij}\}}{\max_{1 \leq i \leq n} \{x_{ij}\} - \min_{1 \leq i \leq n} \{x_{ij}\}} \tag{5}$$

where, y_{ij} is the value of the j evaluation index of the i sample which has been dimensionless, x_{ij} is the original index value of the i index of the j sample.

3.3 Results

3.3.1 Predicting with GM (1, 1) Model

According to the GM (1, 1) model prediction steps [6], the explanatory variable is substituted into MATLAB 2020a version to calculate the prediction results, as shown in Fig. 1.

The prediction results of GM (1, 1) model are shown in Fig. 1. The average relative error of GM (1, 1) prediction value is 1.6143e−4. In this study, the posteriori error method was used to test the suitability of the model, $S1 = 431.7636$, $S2 = 4.2618$, and the posterior error ratio $C = 0.0099$, $P > 0.9$, according to the accuracy level, the prediction accuracy of the GM (1, 1) model is excellent.

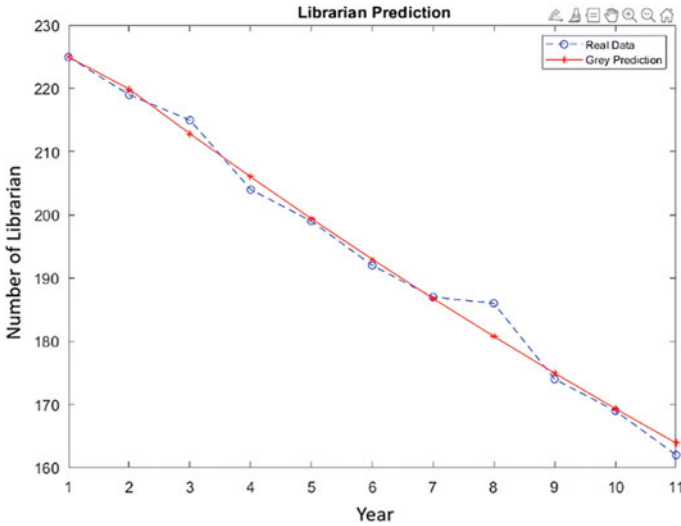


Fig. 1 Number of librarians prediction: GM (1, 1) model

3.3.2 Predicting with BP Neural Network

We import the standardized data into MATLAB 2020a version for neural network training, the number of input layers is 4, the number of hidden layers is 4, and the number of output layer is 1. In this study, sigmoid function is used to calculate the hidden layer and MATLAB purelin function is used to calculate the output layer, and the maximum number of iterations is set to 1000, the target accuracy is set to 1×10^{-4} , training rate is set to 0.1, and the Levenberg–Marquardt (trainlm) is used to training algorithm. This algorithm has faster convergence speed and higher training accuracy than other algorithms. The ratio of training samples to test samples is 3:1, that is, the number of librarians from 2010 to 2017 is the training samples, and the number of librarians from 2018 to 2020 is the test samples. The training results are shown in Fig. 2.

The BP neural network training performance is illustrated in Fig. 2, which shows that when the learning times of BP neural network is 28, the curve begins to converge, and the prediction accuracy is $6.9572e-04$, which meets the preset accuracy requirements.

BP neural network prediction results are shown in Fig. 3, the result shows that the difference between the real data and the predicted value is small, indicating that the training effect is good and the prediction result is reasonable.

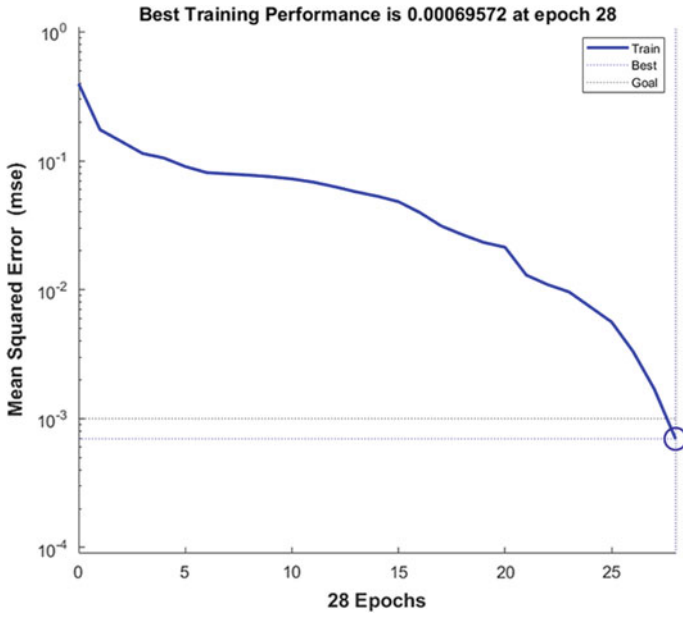


Fig. 2 BP neural network training performance

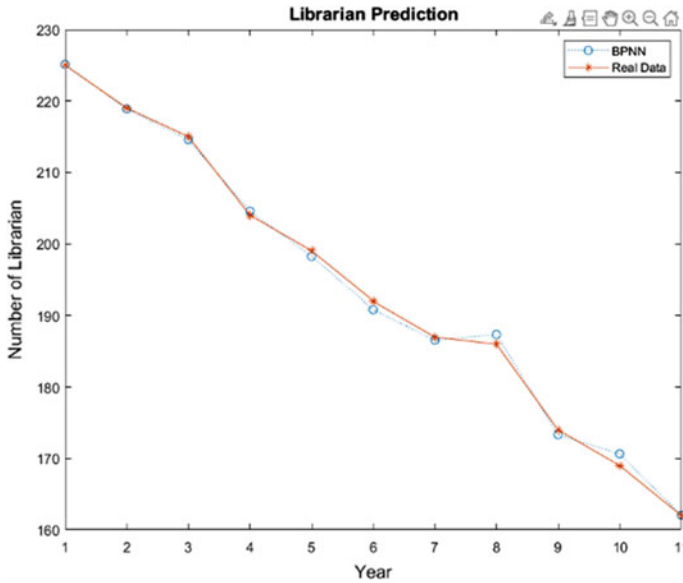


Fig. 3 Number of librarians prediction: BP neural network

Table 2 GM (1, 1)-BP neural network weight table

Model	GM (1, 1)	BPNN
Weight	0.140350877	0.859649123

3.3.3 Predicting with GM (1, 1)-BP Neural Network

According to the construction method of GM (1, 1)-BP neural network combined model, the weight of GM (1, 1)-BP neural network model is calculated, as shown in Table 2.

According to the weight value of GM (1, 1) model and BP neural network in Table 2, a combined prediction model is constructed:

$$y_t = 0.140350877L_{1(t)} + 0.859649123L_{2(t)} \tag{6}$$

We take the data predicted by GM (1, 1) model and BP neural network into formula 6 to calculate the results of the combined model, on this basis, compared with the predicted value calculated by GM (1, 1) model and BP neural network model alone, the results are shown in Table 3.

Table 3 shows that since 2010, the number of librarians of Sichuan University Library has continued to decrease, from 225 in 2010 to 162 in 2020, with a decrease rate of 28%. It can be predicted that the number of librarians will continue to decline in the next few years. The average relative error of GM (1, 1) model fitting value and BP neural network fitting value is 0.71% and 0.38%, respectively. It shows that the prediction accuracy of the two models is high. By analyzing the relative error of a single prediction model, it can be found that GM (1, 1) model has low prediction

Table 3 Comparison of prediction

Year	Real data	Predicting value			Relative error		
		GM (1, 1)	BPNN	GM (1, 1)-BPNN	GM (1, 1) (%)	BPNN (%)	GM (1, 1)-BPNN (%)
2010	225	225	225	225	0.00	0.00	0.0
2011	219	220	219	219	0.46	0.00	-0.06
2012	215	213	214	214	-0.93	-0.47	-0.53
2013	204	206	205	205	0.98	0.49	0.56
2014	199	199	198	198	0.00	-0.50	-0.43
2015	192	193	191	191	0.52	-0.52	-0.37
2016	187	186	186	186	-0.53	-0.53	-0.53
2017	186	180	187	186	-3.23	0.54	0.01
2018	174	175	173	173	0.57	-0.57	-0.41
2019	169	169	170	170	0.00	0.59	0.51
2020	162	163	162	162	0.62	0.00	0.09

accuracy for abnormally fluctuating data, and the prediction accuracy of BP neural network has always remained stable. The average relative error of the fitting value of the combined model is 0.32%, which further improves the prediction accuracy of the model and can be better used to predict the change trend of the librarian.

4 Discussion

In this study, the prediction accuracy of GM (1, 1) model is excellent, according to Yao's research, the growth rate of a sequence is generally constant, while a monotonically decreasing sequence is susceptible to random factors, which will lead to undesirable changes. Therefore, when the data sample is small, the sequence changes are more stable [6]. In this work, the data is a monotonically decreasing small sample data, which can maintain relative stability, so the prediction accuracy is high. However, among the three models, the relative error of GM (1, 1) model is still greater than that of BP neural network and the combined model, and the relative error in 2017 was the largest, which was 3.32%. The possible reason is that GM (1, 1) model is more suitable for linear data, but the change of the number of librarians is affected by many factors. It is a non-linear system, so the prediction accuracy is lower than the other two models. The prediction relative error of the BP neural network is in the range of $[-1\%, 1\%]$, and the prediction accuracy is generally higher than that of the GM (1, 1) model. The purchase cost of literature resources, electronic resources, paper resources and annual expenses included in this study directly affect the workload of the library. The higher the four indicators, the greater the workload, the more librarians are needed. On the contrary, the lower the workload, the less librarians are needed. Similar to Li's research [7], we found that the combined model of GM (1, 1)-BP neural network is better than the single GM (1, 1) model and BP neural network, and the fitting performance is better. Some scholars' research shows that both GM (1, 1) model and BP neural network have a large error when they are used separately [8], since the data in this study is a monotonically decreasing small sample data, and the explained variable is closely related to the explanatory variables, this situation did not occur.

JL Sun proposed that 5% of the value created by the library is generated by the library buildings, 20% by the information resources and 75% by the librarians [9]. The sample data of this study shows that since 2010, the Sichuan University Library has significantly increased its annual expenses and workload, while the number of librarians has been declining. At the same time, with the development of information technology, the work content and working methods of the library have undergone tremendous changes compared with the past. Therefore, librarians must constantly learn new technologies and change from traditional librarians to subject librarians and digital librarians. The library should timely organize relevant training, purchase corresponding equipment, and vigorously strengthen the enthusiasm of librarians for the application of information technology [10]. We should also note that due to the continuous influx of new technologies and knowledge, the increasing requirements

for the ability of librarians, coupled with factors such as age, marriage, and readers, have brought pressure to librarians. Once this pressure continues for a long time, librarians will gradually enter a state of job burnout, which is not conducive to the development of library work [11].

There are some limitations in this study. Firstly, this study collected sample data of Sichuan University Library for only 11 years from 2010 to 2020, and the sample size is relatively small. Secondly, methods for predicting the evolution trend of time sequence data include multiple regression models, autoregressive integrated moving average model (ARIMA), etc., and this study only uses GM (1, 1) model and BP neural network. Thirdly, due to the availability of data, the explanatory variables of this study only include the purchase cost of literature resources, electronic resources, paper resources and annual expenses. Therefore, in the next step of research, scholars can use a variety of prediction methods to compare the results, and add more explanatory variables to study more cases to make the prediction results more effective.

5 Conclusion

This research constructs a GM (1, 1)-BP neural network combined model, and takes Sichuan University Library as a case study to conduct an empirical analysis. Results show that, the prediction accuracy of the GM (1, 1)-BP neural network combined model is better than that of a single prediction model. From 2010 to 2020, the number of librarians continued to decline, the library management department should pay attention to this trend and take practical measures to actively respond to the adverse effects of the reduction of librarians to promote the development of the library.

References

1. Chen, K., Jiang, X., Huang, H., et al.: An analysis of research on human resource management in university library. *J. Library Sci. Soc. Sichuan* **03**, 9–13 (2017)
2. Pinfield, S.: The changing role of subject librarians in academic libraries. *J. Librariansh. Inf. Sci.* **33**(1), 32–28 (2001)
3. Sarasvathy, P., Nambratha, G.R., Giddaiah, D.: Changing roles of the librarians in the virtual/digital era. *SRELS J. Inf. Manag.* **49**(5), 495–500 (2019)
4. Steering Committee for Academic Libraries of China. Statistical Data. [EB/OL]: <http://www.scal.edu.cn/tjpg/tjsj>. Accessed on 25 Sept 2021
5. Wu, J.L.: Study on GM(1,1)-BP combination model based on grey correlation. *J Chongqing Univ Technol (Nat Sci)* **33**(11), 207–210 (2019)
6. Yao, T.X., Liu, S.F., Xie, N.M.: On the properties of small sample of GM (1, 1) model. *Appl. Math. Model.* **33**(4), 1894–1903 (2009)
7. Li, M.H., Zhao, M.G., Liu, M., et al.: Energy consumption prediction of campus building based on the GM-BP neural network. *Build Energy Eff* **44**(11), 87–90 (2016)
8. Guo, D.J., Zou, Y.: Prediction of China's urban residents consumption level based on Grey GM (1, 1)-BP neural network. *Adv. Serv Sci. Serv. Inf. Technol.* **52**, 193–199 (2014)

9. Sun, J.L.: The importance of human resource management in library reform. *Library Tribune* **22**(5), 133–135 (2002)
10. Ramzan, M., Asif, M., Ahmad, S.: Librarians' attitudes towards application of information technology in academic libraries in Pakistan. *Inf. Res. Int. Electron. J.* **26**(1), 887 (2021)
11. Smith, D.L., Bazalar, B., Wheeler, M.: Public librarian job stressors and burnout predictors. *J. Libr. Adm.* **60**(4), 412–429 (2020)