Research on the Funds Forecasting and Quality Control of Book Purchasing

Shilian Cai[*]

The University Library**,** Beijing University of Civil Engineering and Architecture**,** Beijing 102616, P. R. China
*Corresponding author: caishilian@163.com

Abstract It is of great significance for optimizing the allocation of book purchasing funds and guaranteeing the quality of books procured improve the utilization efficiency of library resources as well as the quality of literature service. Based on the circulation data in the library at Beijing University of Civil Engineering and Architecture in the past 10 years, this paper studies the allocation of funds and quality control of books purchased using a neural network model. Firstly, a prediction model with the BP neural network is established based on the circulation data of books in the past to provide necessary reference data for the budget of the book purchasing funds. Then, an RBF neural network model is used to predict the allocation of purchasing funds. Finally, a book purchasing quality analysis system is established through the study of the matching degree between the average professional books of students in each discipline and the average professional books of the students and the correlation coefficient between purchasing and borrowing. This paper provides a clear criterion for judging whether book purchases meet the needs of disciplines and readers' requirements. It also provides a scientific basis for the university library to make a reasonable plan on book purchases.

Keywords: books, circulation, purchase funds, neural network, purchase quality

Cite This Article: Shilian Cai, "Research on the Funds Forecasting and Quality Control of Book Purchasing." *Journal of Computer Sciences and Applications*, vol. 6, no. 1 (2018): 23-31. doi: 10.12691/jcsa-6-1-3.

1. Introduction

Commonly used prediction methods include moving smoothing method, exponential smoothing method [1,2], regression analysis method [3], grey prediction method [4], and combination of multiple forecasting methods. Most of these methods focus on the analysis of causality regression models or time series models. The established models cannot reflect the inherent structure and complex characteristics of the predicted dynamic data scientifically and essentially, without some information loss. As a parallel computing model, artificial neural network has many advantages over traditional modeling methods: 1) it has a good ability for nonlinear mapping; 2) it does not require much prior knowledge of the object modeled. If the input and output data of the object are given, the nonlinear relationship between the input and the output can be modelled through the learning process of the network itself.

The circulation of books is one of key factors to measure the number of books used by readers and the service levels of a library. The library staff can use the circulation of books in time to adjust the structure of the collection, improve the utilization efficiency of the collection, and to improve the service approaches and service quality of the library. However, the circulation of books is influenced by many factors such as the quality of books, the time of the library's opening, the setting of disciplines and specialties, the service quality of librarians, and the level of management, etc. This situation shows a highly nonlinear relationship. It is difficult to find a suitable traditional method for an effective prediction. In the literature [5], one researcher adopted the method of combining an empirical model with a mathematical method to determine the distribution of collected copies of books within a certain period, but this method is often limited by the level of understanding of the procurers. A statistical regression analysis was conducted to create a single regression model between the ratio of readers' refusing to borrow a book and the copy quantity of the book, which fulfills the control of the ratio of refusing to borrow the book and the prediction of the number of coping the book [6]. The quantity of duplicate purchases of a certain kind of book was synthetically determined by considering the combined effect of seven main factors on the volume of duplicates [7]. Based on multiple regression analysis theory, the linear correlation between circulation rate, borrow rejection rate, and average copy quantity was analyzed; the prediction method of the volume of books in a library was discussed using actual data [8]. The budget and allocation of funds are very complicated. There are many factors affecting the expenditure of literature funds in a library; their interactions are complex; and their relationships are nonlinear. A grey system theory was used to predict the purchasing funds of a library's literatures [9]. The quality analysis of book procurement is an important part of library procurement; the quality of books purchased directly affects the quality and utilization of books. The quality of books purchased was analyzed by means of an explanatory structure model and the library distribution [10,11,12].

Based on the historical data of the library at Beijing University of Civil Engineering and Architecture, this paper performs a numerical simulation of the circulation of books and the cost of purchasing books using the BP and RBF neural network. In addition, the quality of book purchases is tested by the quantitative index of the matching degree of the per capita share of books for each specialty and discipline, and the correlation coefficient between purchasing and borrowing.

2. Book Circulation Prediction

2.1. Data Preprocessing and Neural Network Structure

Based on the data (as shown in Table 1) of the circulation of books from 2002 to 2010 in the Library at Beijing University of Civil Engineering and Architecture, a specific neural network prediction model is established.

Table 1. Book Circulation (2002-2010)

Year	Book Circulation (volume)	Year	Book Circulation (volume)
2002	45761	2007	77519
2003	60345	2008	64551
2004	68588	2009	62345
2005	79885	2010	63508
2006	85260		

First, the data is normalized, and the processing formula is as follows [13]:

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \qquad (1)$$

x_i as raw data; $x_{\max}, x_{\min}$ is the maximum and $x_{\max}$ and $x_{\min}$ are the maximum and the minimum value of the raw data, respectively. x_i' *is* normalized data. Table 2 shows the normalization results of the data in Table 1.

Table 2. The Data from 2002 to 2010 after a Normalized Treatment

Year	Book Circulation
2002	0.0000
2003	0.3697
2004	0.5787
2005	0.8651
2006	1.0000
2007	0.8051
2008	0.4764
2009	0.4204
2010	0.4499

2.2. BP Neural Network Method

2.2.1. Neural Network Structure

The neural network mainly includes the number of network layers, the number of neurons, the activation functions in each layer, and the neural network structural diagram [14] as shown in Figure 1.
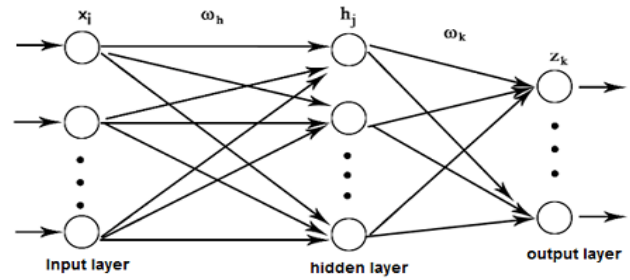

Figure 1. BP Neural Network Structure Diagram

2.2.2. Sample Design

Book circulation is a series of numbers arranged in a chronological order, as in Table 3.

Table 3. Segmentation Method for Sample Data

N Inputs				M Outputs			
x_1	x_2	$\cdots$	x_N	x_{N+1}	x_{N+2}	$\cdots$	x_{N+M}
x_2	x_3	$\cdots$	x_{N+1}	x_{N+2}	x_{N+3}	$\cdots$	x_{N+M+1}
	……				……		
x_K	x_{K+1}	$\cdots$	x_{N+K-1}	x_{N+K}	x_{N+K+1}	$\cdots$	$x_{N+M+K-1}$

Matlab software is used in this paper, A BP neural network with three layers is used to establish a forecast model for book circulation. The data from 2002 to 2008 are selected as training sample data and data from 2009 to 2010 are selected as testing sample data. These data samples are shown in Table 4. The transfer functions of the hidden layer and output layer are "logsig" and "purelin", respectively; and the training function is "trainlm".

Table 4. Sample Data

No. of Training Sample	Input	Output
1	Circulation from 2002 to 2004	Circulation in 2005
2	Circulation from 2003 to 2005	Circulation in 2006
3	Circulation from 2004 to 2006	Circulation in 2007
4	Circulation from 2005 to 2007	Circulation in 2008
No. of Testing Sample	Input	Output
1	Circulation from 2005 to 2008	Circulation in 2009
2	Circulation from 2007 to 2009	Circulation in 2010

2.2.3. Sample Data Training and Results Prediction

The data samples from 2002 to 2007 are used to train the input network using Matlab7.0 software. The number of hidden nodes is trained from 3 to 12, and the predicted result is slightly better when the number is 5. The allowable error is 0.0001; the training is completed after two iterations; and the error is below the error tolerance. Network errors during training are shown in Figure 2.
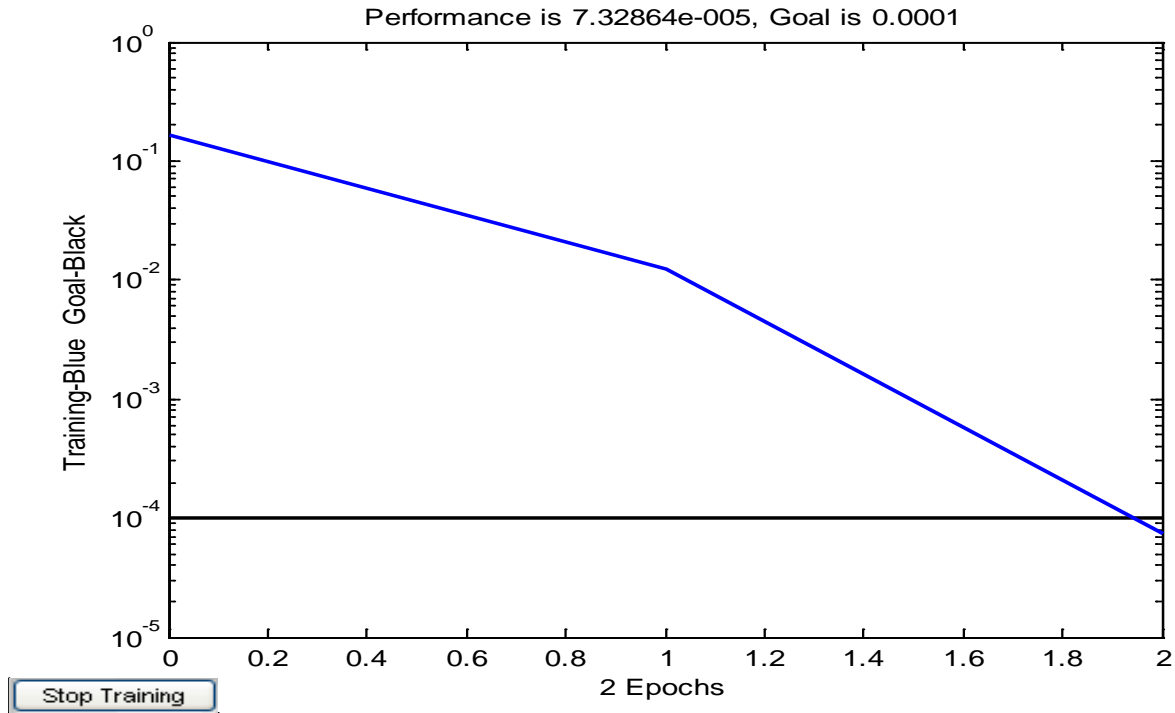
Figure 2. The Curve of Errors during Training

Table 5. Comparison of Expected Output and Actual Output

Year	Expected Output	Actual Output	Prediction Error
2005	0.8639	0.8639	0
2006	1.0000	1.0000	0
2007	0.8040	0.8040	0
2008	0.4757	0.4757	0
2009	0.4204	0.3836	0.0368
2010	0.4499	0.5055	-0.0556

The circulation from 2002 to 2004 are used as training data; the data in 2005 are used as simulation data; and the circulation is predicted by the neural network. The circulation from 2003 to 2005 are used as training data; the data in 2006 are used as simulation data; and the circulation is predicted by the neural network. The circulation from 2004 to 2006 are used as training data; the data in 2007 are used as simulation data; and the circulation is predicted by the neural network, and so on. Results are shown in Table 5.

2.2.4. Simulation Result and Analysis

The last two rows in Table 5 are testing samples and the other rows are training samples. It is shown that the network outputs are in good agreement with the expected responses for training samples and the errors of predicted values of the testing samples are 0.0368 and -0.0556, respectively. The relative errors of the two testing samples are small, which indicates that this prediction is successful. For plotting the results, the following command is entered in the Matlab command line:

plot([2005:2010], p0(4:9), '-r', [2005: 2010], result, '-o')*

The result is shown in Figure 3, where the curve marked with "*" represents the actual values; the curve marked with "o" represents the predicted values; and the last three points are the prediction of testing sample points.
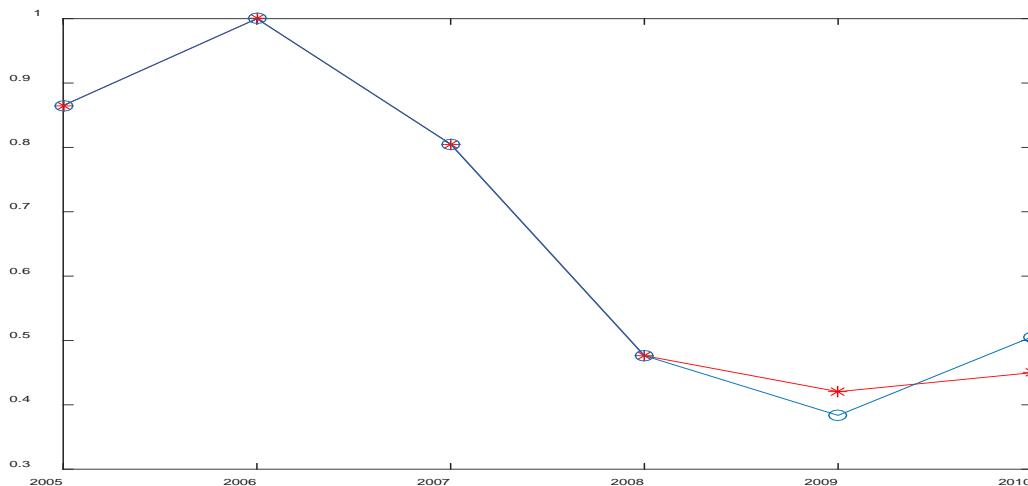

Figure 3. Predicted Results and Real Data

As can be seen from Figure 3, the two curves fit well from 2005 to 2008, which are the results of training, samples and then the forecasted values deviate gradually from the real values. The prediction errors in 2009-2010 gradually increase, but the errors are less than 6%, which shows that the prediction accuracy of the network is acceptable. If a larger sample is used, a better prediction accuracy may be obtained, which is an issue that needs further study.

3. Forecast of Book Purchasing Funds

Taking the library at Beijing University of Civil Engineering and Architecture from 2000 to 2012 as the observation period (as shown in Table 13), a specific neural network prediction model is established.

3.1. *RBF* Neural Network Model Structure and Data Processing

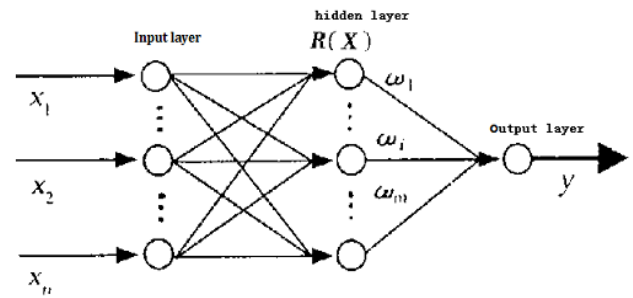The *RBF* network structure used in this paper is shown in Figure 7 [15].

Figure 4. RBF Network Structure

Based on the statistical data of the library circulation system in the past years (2000-2012), five main factors affecting the expenditure of literatures are selected: the number of undergraduate students, the number of graduate students, the number of faculty, the volume of circulation and the volume of books (see Table 6). The expenditure on library literatures (including books, periodicals, and databases) are shown in Table 8. The data in Table 6 are normalized by formula (1) to better use the artificial neural network model and Table 7 is obtained.

Table 6. Readership, Circulation and Volume of Books in 2000-2012

Year	Number of Graduate Students	Number of undergraduates	Circulation (volume)	Number of Faculty	Volume of Books (ten thousand copies)
2000	50	950	98000	150	28
2001	58	1007	101211	170	30
2002	64	1336	141311	183	35
2003	101	1476	201855	201	37
2004	103	1380	217893	263	39
2005	143	1037	253959	340	50
2006	215	1041	299526	425	60
2007	205	1038	273343	402	60
2008	260	1512	238606	429	64
2009	300	1510	196627	463	69
2010	362	1510	188134	446	80
2011	385	1610	170312	567	80
2012	430	1720	170111	600	120

Table 7. Normalized Data in 2000-2012

Year	Number of Graduate Students	Number of undergraduates	Circulation (volume)	Number of Faculty	Volume of Books (ten thousand copies)
2000	0	0	0	0	0
2001	0.021	0.074	0.0159	0.0444	0.0217
2002	0.0368	0.5013	0.2149	0.0733	0.0761
2003	0.1342	0.6831	0.5153	0.1133	0.0978
2004	0.1395	0.5584	0.5949	0.2511	0.1196
2005	0.2447	0.113	0.7739	0.4222	0.2391
2006	0.4342	0.1182	1	0.6111	0.3478
2007	0.4079	0.1143	0.8701	0.56	0.3478
2008	0.5526	0.73	0.6977	0.62	0.3913
2009	0.6579	0.7273	0.4894	0.6956	0.4457
2010	0.8211	0.7273	0.4473	0.6578	0.5652
2011	0.8816	0.8571	0.3588	0.9267	0.5652
2012	1	1	0.3578	1	1

Table 8. Funds for Books, Periodicals and Databases in 2000-2012

Year	Book Purchase Funds	Normalized	Databases Procurement Funds	Normalized	Periodical Purchasing Funds	Normalized
2000	14	0	75	0	15	0
2001	17	0.013	78	0.0205	17	0.0303
2002	35	0.092	80	0.0342	34	0.2879
2003	59	0.197	95	0.1369	32	0.2676
2004	83	0.301	100	0.1712	42	0.4091
2005	170	0.681	221	1	43	0.4242
2006	113	0.432	212	0.9383	44	0.4394
2007	42	0.122	112	0.2534	44	0.4394
2008	66	0.227	119	0.3013	44	0.4394
2009	122	0.472	122	0.3219	50	0.5303
2010	128	0.498	171	0.6575	53	0.5758
2011	217	0.886	211	0.9315	55	0.6061
2012	243	1	216	0.9658	81	1

3.2. The Training and Testing of *RBF* Neural Network Model

The data from 2000 to 2010 are selected as the training dataset, that is, the influencing factors of library literatures funds from 2000 to 2010 are utilized as the training input and the literatures funds from 2000 to 2010 are used as the training output, and the data from 2011 to 2012 are applied as the testing dataset. Therefore, the number of neurons in the input layer is 5, the number of neurons in the output layer is 3, and the number of samples is 13. The network creation code is:

$$net = newrbe(P, T, spread)$$

where *spread* is the distribution density of radial basis functions; *P* and *T* are the input vector and output vector in the training samples, respectively. *newrbe* can create an accurate *RBF* neural network, specifically, the network creation process is a training process. The first 11 samples are used as training samples and the last 2 samples as testing samples. The testing code is:

$$Y = sim(net, P\_test)$$

where P_test is the testing sample vector. Thus, the predicted values are obtained and shown in Table 9.

Table 9. Predicted Results

Year	Book Prediction Value	Real Value of Books	Absolute Error	Relative Error
2011	0.7758	0.8860	-0.1102	-12%
2012	0.7695	1	-0.2305	-23%
Year	**Database Prediction Value**	**Database Actual Value**	**Absolute Error**	**Relative Error**
2011	0.8390	0.9315	-0.0925	-9.9%
2012	0.8689	0.9658	-0.0969	-10%
Year	**Periodical Prediction Value**	**Periodical Actual Value**	**Absolute Error**	**Relative Error**
2011	0.5873	0.6061	-0.0188	3.1%
2012	0.7182	1	-0.2818	28%

Because the training sample size is relatively small, the prediction accuracy is not very high. Due to this factor, the forecast in this paper is acceptable. The forecast results can be used to guide practical work.

4. Purchasing Quality Analysis

4.1. Standard Deviation Analysis of Book Purchasing Quality

4.1.1. Distribution of the Volume of Books for Each Student in Various Disciplines

Using one year as the time for statistics, the total number of books purchased for a year is *U*, and the total number of students majoring in a discipline is *R*. The average annual number of books for students $\overline{X}$ is:

$$\overline{X} = U/R \text{ (a constant for a given year)} \quad (2)$$

If there are U_i books in category *i* and R_i students in category *j*, then the average number of professional books per year for that subject Y_i is:

$$Y_i = U_i/R_i \ (i = 1, 2, 3, \cdots n). \quad (3)$$

4.1.2. Evaluation Index of Purchasing Quality Based on Standard Deviation

Standard deviation is a standard to measure the degree of dispersion of data distribution, to be precise, it represents the degree of dispersion between data in a set of multiple measurements, and measures the degree of deviation from the arithmetic mean value. The smaller the standard deviation, the less these values deviate from the average. In this paper, the matching degree σ between the average possession of students in each subject and the total average ownership of books can be expressed by the following formula:

$$\sigma = \sqrt{\frac{1}{n}\sum_{1}^{n}(Y_i - \overline{X})^2} \quad (4)$$

The smaller the value of σ, the closer the relationship between Y_i and $\overline{X}$. The more similar the two, the higher the quality of books purchased. On the contrary, the quality of books procured is lower.

4.1.3. Application Examples

Books procured for science and engineering in 2013 and 2014 in the library at Beijing University of Civil Engineering and Architecture and the quality of books purchased are analyzed using the above model.

(1) Book purchasing situation

The distributions of the disciplines in Humanity and Law, Economics, Sciences, Geomatics and Urban Spatial Informatics, Mechanical and Electronics, Telecommunication, Architectural and Civil Engineering, and Environment and Energy at the university in 2013 and 2014 are shown in Table 10.

(2) Distribution of the average number of books for students in each discipline

The number of students in the above eight disciplines is taken as the statistical object to observe the average amount of specialized books per student. The distribution of the average number of books per student in 2013 and 2014 is shown in Table 11.

(3) Calculated result and analysis

The matching degree σ between the per capita possession of specialized books in each discipline and the total per capita possession of specialized books in 2013 and 2014 is calculated using formula (4) as follows:

$$\sigma_{2013} = \sqrt{\frac{1}{n}\sum_1^n (Y_i - \overline{X})^2} \approx 29.29,$$

$$\sigma_{2014} = \sqrt{\frac{1}{n}\sum_1^n (Y_i - \overline{X})^2} \approx 22.88$$

Because of $\sigma_{2014} < \sigma_{2013}$, compared with 2013, the per capita professional book possession in 2014 is close to the total per capita professional book possession, indicating that the quality of purchases in 2014 is slightly higher than that in 2013. The main reason is that class P and X increased when the total volume of book purchases declined in 2014.

4.2. Analysis of the Quality of Books Purchased Based on the Correlation Coefficient

The quantity of books borrowed is a statistical index to measure the quality of books purchased. Statistics and analysis for the quantity of books borrowed on a regular basis can help the staff understand the circulation of books in time and the characteristics of readers' demands for books, which will help improve the quality of book purchasing. The database of Chinese paper document collection and the utilization in the library at Beijing University of Civil Engineering and Architecture is classified and input by MELINETS II Library Automation system according to "Chinese Library Classification". Based on the database, the utilization of professional books in 2013 and 2014, the corresponding professional book purchases in 2013 and 2014 are counted. A mathematical model is used to analyze the quality of book purchasing using the data from these two years.

Table 10. Distributions of Subject Structure of Books Purchased in 2013and 2014

Category		C, D	F	O	P	TB\TF\TG\TH	TM\TN\TP	TU	X	Total
	i	1	2	3	4	5	6	7	8	
2013	Types	7022	8197	2024	558	2270	6252	9916	870	37109
	Number of books	12483	14985	4194	992	4047	11623	17066	1520	66910
2014	Types	5239	6208	1369	508	1940	4840	7157	1074	28335
	Number of books	10330	12384	2831	1015	3910	10083	14714	2115	57382
	Related Colleges	Humanity and Law	Economics	Sciences	Geomatics and Urban Spatial Informatics	Mechanical and Electronics	Telecommuni cation	Architecture and Civil	Environment and Energy	

Table 11. The Distribution of the Average Number of Specialized Books for Students in 2013-2014

Category		C,D	F	O	P	TB\TF\TG\TH	TM\TN\TP	TU	X	Total
	i	1	2	3	4	5	6	7	8	
U_i	2013	12483	14985	4194	992	4047	11623	17066	1520	66910
	2014	10330	12384	2831	1015	3910	10083	14714	2115	57382
R_i	2013	128	210	110	128	160	232	484	258	1710
	2014	128	237	104	132	160	222	507	260	1750
Y_i	2013	97.52	71.36	38.13	7.75	25.29	50.10	35.26	5.89	39.13
	2014	80.70	52.25	27.22	7.69	24.44	45.42	29.02	8.13	32.79

4.2.1. Distribution of Professional Structure of Book Purchasing

Using the period of one year for statistics, the total number of books purchased for a year is U, of which category i has a volume U_i, the frequency X_i of which is as follows:

$$X_i = (U_i/U) \times 100\% \quad (\text{where } U = \sum_1^n U_i). \quad (5)$$

The purchase quantity of each specialty is not equal. There are always a few professional books with a larger frequency and other professional books with a smaller frequency. Selecting the total number of specialties n, the frequency can form a purchasing frequency distribution vector (discrete distribution function).

$$\overrightarrow{X} = (X_1 \quad X_2 \quad \cdots \quad X_n)^T \quad (6)$$

4.2.2. The Distribution of Specialized Structure in Book Circulation

The categories with the same numbers in the professional structure of book procurement are investigated. Because there is a time difference between purchase and circulation, the circulation is counted by the time of circulation falling behind the time of purchase for one year. Within the statistical period, the total number of books in circulation of Category i is U_i', the total number of books in circulation within the statistical period is U'. The frequency of circulation of the total number of books of all kinds is as follows:

$$Y_i = (U_i'/U') \times 100\% \quad (\text{among } U' = \sum_1^n U_i'). \quad (7)$$

According to the circulating frequency Y_i, the distribution vector of circulation frequency (discrete distribution function) can also be computed:

$$\overrightarrow{Y} = (Y_1 \quad Y_2 \quad \cdots \quad Y_4)^T. \quad (8)$$

Vectors $\overrightarrow{X}$ and $\overrightarrow{Y}$ have the following characteristics:
(1) The sum of frequencies is 1, namely

$$\sum_{i=1}^n X_i = 1; \sum_{i=1}^n Y_i = 1.$$

(2) The mean value of two vectors is equal, namely

$$\overline{X} = 1/n; \overline{Y} = 1/n.$$

4.2.3. Distribution Vector Preprocessing

Standardized preprocessing of $\overrightarrow{X}$ and $\overrightarrow{Y}$ is carried out according to the following formula, respectively:

$$X_i' = \frac{X_i - \overline{X}}{S_x}; Y_i' = \frac{Y_i - \overline{Y}}{S_y}. \quad (9)$$

where S_x and S_y are the standard deviations of distribution X and Y, respectively:

$$S_x{}^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \overline{X})^2$$

$$S_y{}^2 = \frac{1}{n-1} \sum_{i=1}^n (Y_i - \overline{Y})^2 \quad (10)$$

where

$$\overrightarrow{X'} = (X_1' \quad X_2' \quad \cdots \quad X_n'), \overrightarrow{Y'} = (Y_1' \quad Y_2' \quad \cdots \quad Y_n').$$

The two vectors are transformed into zero mean and standardized vectors of unit variance, balancing the influence of each major.

4.2.4. Correlation coefficient r between Book Purchasing and Circulation

Mathematically, the similarity of two discrete distribution vectors $\vec{a}$, $\vec{b}$ can be measured by the cosine of the angle between the two vectors, i.e.:

$$COS\theta = \frac{\vec{a} \bullet \vec{b}}{|a||b|}.$$

As a result, the degree of similarity between the professional structure distribution vector $\overrightarrow{X}$ of book procurement and the distribution vector $\overrightarrow{Y}$ of the professional structure in circulation can be expressed in the following terms:

$$\cos = \frac{\overrightarrow{X} \bullet \overrightarrow{Y}}{|X||Y|} = \frac{\sum_{i=1}^n X_i Y_i}{\left[\left(\sum_{i=1}^n X_i^2 \right) \left(\sum_{i=1}^n Y_i^2 \right) \right]^{1/2}} \quad (11)$$

If $\overrightarrow{X}$ and $\overrightarrow{Y}$ are normalized preprocessed as $\overrightarrow{X'}$ and $\overrightarrow{Y'}$, then the angle cosine between $\overrightarrow{X'}$ and $\overrightarrow{Y'}$ is their correlation coefficient:

$$r = \cos\theta = \frac{\sum_{i=1}^n X_i' Y_i'}{\left[\left(\sum_{i=1}^n X_i'^2 \right) \left(\sum_{i=1}^n Y_i'^2 \right) \right]}. \quad (12)$$

Because $0 \le \cos\theta \le 1$, when $|\cos\theta|(|r|) \to 1$, $\theta \to 0°$.

The closer the relationship between $\overrightarrow{X}$ and $\overrightarrow{Y}$ is, the more similar the two are; therefore, book purchasing is in line with circulation. Conversely, if $|r| \to 0$, then $\theta \to 90°$, the two are dissimilar. Therefore, we use the correlation coefficient of the book purchase and circulation to evaluate the quality of book purchases.

4.2.5. Examples in Application

(1) Distribution of book purchase and circulation in the library at Beijing University of Civil Engineering and Architecture in 2013

Table 12. Frequency Distribution of the Structure of Professional Books in 2013

Category	D	F	O	TB	TH	TM	TN	TP	TU
i	1	2	3	4	5	6	7	8	9
U_i	9934	14985	4194	1695	1560	1494	2436	7693	17066
X_i	0.1627	0.2454	0.0687	0.0278	0.0255	0.0245	0.0399	0.1260	0.2795
$X_i - \bar{X}$	0.0516	0.1343	-0.0424	-0.0833	-0.0856	-0.0866	-0.0712	0.0149	0.1684
$X_i^{'}$	0.5239	1.3636	-0.4305	-0.8458	-0.8691	-0.8793	-0.7229	0.1513	1.7098
$X_i^{'2}$	0.2745	1.8594	0.1853	0.7154	0.7554	0.7731	0.5226	0.0229	2.9234

Table 13. Distribution of the Frequency of Book Circulation in 2013

Category	D	F	O	TB	TH	TM	TN	TP	TU
i	1	2	3	4	5	6	7	8	9
$U_i^{'}$	3014	8390	11638	2983	1546	1974	1793	17254	65323
Y_i	0.0265	0.0737	0.1022	0.0262	0.0136	0.0173	0.0157	0.1515	0.5734
$Y_i - \bar{Y}$	-0.0846	-0.0374	-0.0089	-0.0849	-0.0975	-0.0938	-0.0954	0.0404	0.4623
$Y_i^{'}$	-0.4705	-0.2080	-0.0495	-0.4722	-0.5423	-0.5217	-0.5306	0.2247	2.5712
$Y_i^{'2}$	0.2214	0.0433	0.0025	0.2230	0.2941	0.2722	0.2815	0.0505	6.6110

Considering nine kinds of professional books in 2013, including Politics and Law (D), Economics (F), Mathematics/Physics/Chemistry (O), Industries and Technologies (TB), Mechanical Instruments (TH), Electrical Technology (TM), Wireless Technology and Telecommunications (TN), Automation and Computer Science (TP), and Architectural Engineering (TU). Distributions of the professional structure and the circulation are shown in Table 12 and Table 13.

$$\sum_{i=1}^{9} X_i^{'2} = 8.0320, \sum_{i=1}^{9} Y_i^{'2} = 7.9994, \sum_{i=1}^{9} X_i^{'} Y_i^{'} = 5.6344.$$

Based on the formula (12), the correlation coefficient of book procurement and circulation in 2013 is calculated and the result is $r = 0.7029$. It is shown that the books purchased are positively correlated with the circulation data, which indicates that the books utilization ratio has a profound influence on the purchase.

(2) Distribution of books purchased and circulation in the library at Beijing University of Civil Engineering and Architecture in 2014

Considering the same nine categories of professional books mentioned above, the frequency distribution of the structure and circulation of specialized books in 2014 is shown in Table 14 and Table 15:

Table 14. Frequency Distribution of the Structure of Professional Books in 2014

Category	D	F	O	TB	TH	TM	TN	TP	TU
i	1	2	3	4	5	6	7	8	9
U_i	7720	12384	2831	1698	1414	1846	2004	6233	14714
X_i	0.1518	0.2436	0.0557	0.0334	0.0278	0.0363	0.0394	0.1226	0.2894
$X_i - \bar{X}$	0.0407	0.1325	-0.0554	-0.0777	-0.0833	-0.0748	-0.0717	0.0115	0.1783
$X_i^{'}$	0.4132	1.3452	-0.5624	-0.7888	-0.8457	-0.7594	-0.7279	0.1168	1.8102
$X_i^{'2}$	0.1707	1.8096	0.3163	0.6222	0.7152	0.5767	0.5298	0.0136	3.2768

Table 15. Distribution of the Frequency of Book Circulation in 2014

Category	D	F	O	TB	TH	TM	TN	TP	TU
i	1	2	3	4	5	6	7	8	9
$U_i^{'}$	2071	4906	9048	1902	1014	1020	853	9356	32978
Y_i	0.0328	0.0777	0.1433	0.0301	0.0161	0.0162	0.0135	0.1482	0.5222
$Y_i - \bar{Y}$	-0.0783	-0.0334	0.0322	-0.0810	-0.0950	-0.0949	-0.0976	0.0371	0.4111
$Y_i^{'}$	-0.4801	-0.2048	0.1974	-0.4966	-0.5825	-0.5819	-0.5984	0.2275	2.5205
$Y_i^{'2}$	0.2305	0.0419	0.0390	0.2466	0.3393	0.3386	0.3581	0.0517	6.3530

$$\sum_{i=1}^{9} X_i'^2 = 8.0309, \sum_{i=1}^{9} Y_i'^2 = 7.9987, \sum_{i=1}^{9} X_i' Y_i' = 5.6661.$$

Based on the formula (12), the correlation coefficient of book purchases and circulation in 2014 is calculated and the result is $r = 0.7194$. It is shown that the book purchasing is positively correlated with circulation data, which indicates that the book utilization ratio has a major influence on purchasing. The correlation coefficient in 2014 is slightly higher than that of 2013, which indicates that the quality of book procurement in 2014 is higher than that in 2013.

4.3. Quality Control Analysis

First, the root mean square deviation (RMS), as a measure of the error of the measured values, indicates the degree of dispersion of the data in a set of data obtained by multiple measurements. The smaller the σ value is, the closer the relationship between Y_i and X is; the more similar they are, that is, the closer the per capita specialized books are to the total per capita specialized books, the higher the purchasing quality of books. On the contrary, the greater the σ value, the more estranged the relationship between Y_i and X; the more inconsistent they are, the greater the difference between the per capita occupation of specialized books and the total per capita specialized book, therefore the lower the purchasing quality of books. Because the differences between the 2013 and 2014 data are large, the root mean square deviation σ is also large. Although the quality of procurement in 2014 was slightly higher than in 2013, the quality of procurement in 2013 and 2014 was not high, mainly because the volume of literature resources purchased in the two years is highly imbalanced by category. The P and X categories account for a small proportion. The reason is that the winning bookseller is not in proportion to the professional requirements of books in the university library. Faculty who were interviewed only selected the bibliography given by the bookseller, and purchasers neglected the dynamic situations of the publishing business.

Second, taking the library data and interview data in 2013 and 2014 as an example, the correlation coefficient between the two is greater than 0.7, indicating that the two sets of data are positively related. The use of books has a greater impact on literature procurement, indicating that the more books used the larger number of documents procured. The closer the correlation coefficient is to 1, the more similar the two are. The book purchase is in line with the circulation, so the quality of 2104 procurement is slightly higher than that of 2013. The main reason why the purchasing quality is not ideal is that the readers' demands were not well understood, and purchasers neglected the changes of the units' demands and the balance of the book purchase quantity among all disciplines.

5. Conclusion

From the computation results, a BP neural network can be used to predict the circulation of books and the RBF neural network model works well in predicting the cost of literature procurement; the prediction errors are within an acceptable range, which provides a scientific basis for a budget for procurement funds in the future. In addition, the quality of books purchased, and control criterion were analyzed using situations of readers' using books and the demands of borrowing, which is convenient for purchasers to use in improving their work in literature acquisition at any time.

Acknowledgements

References

[1] Xu Ying. The prediction of the volume and circulation of the Chinese Library [J], journal of intelligence, 1999, 18(6): 53-56.

[2] Lu Xiaobin. Information analysis [M]. Beijing: science and Technology Literature Press, 2007.

[3] Sun Bao, Zhao Yanmei, Wang Zili, Li Jianhui. Research on Procurement quantity Modeling based on Borrowing Statistics [J]. Information science, 2011 / 29: 103-107.

[4] Tan Wenhua. Grey Prediction of Book flow in University University Library [J]. Modern Intelligence: 10: 119-120.

[5] Cao Zhen. The configuration of the Chinese copy of the University Library [J]. Journal of the university library 3: 53-56

[6] Wang Juping. A preliminary study on quantitative Analysis and Prediction methods in Library Science and Information Science [J]. Journal of Intelligence 2007 / 1: 105-106.

[7] Liu Xinwen. Quantitative analysis of the quantity of duplicated books in library [J]. Journal of Southwest China Normal University, 2007(4): 87-89.

[8] Wang Jian. Multiple regression analysis method to predict the average duplicate quantity [J]. Journal of Ningxia Teachers College (social science), 2010: 5: 167-168.

[9] Wang Tingman, Shen Si. Using Grey system Theory to predict the expense of Books and Literature [J]. Journal of Xi'an Institute of Science and Technology, 2002, 22(9): 367-370.

[10] Shi Yanfen. Application of explanatory structure Model in quality Analysis and Control of Book Purchasing [J]. Library work and Study, 2012, 56 (5): 94-97.

[11] Zhu Chunhui. Distribution of Book Utilization and its influence on Collection Construction[J]. Library Development, 2011(3): 56-64.

[12] Wang Fang. An Analytical Model of specialized Books quality in Colleges and Universities and its Application [J]. The Library Journal of Henan, 2007, 27 (1): 10-13.

[13] Zhang Xuelian, Li Liyan. BP Neural Network Forecasting Model of Book and Literature funding [J]. Journal of Intelligence, 2006 (6): 115 -116.

[14] Bai Xuebing. Research on Economic Prediction Model based on BP and RBF Neural Network [J]. Computer knowledge and Technology, 2011, 7(10): 2345-2347.

[15] Zhu Kai, Wang Zhenglin. Proficient in MATLAB neural network [M]. Beijing: Publishing House of Electronics Industry, 2010.