

# Prediction and Estimation of Book Borrowing in the Library: Machine Learning

Jinbao Sun

Graphic Center, Henan Mechanical and Electrical Vocation College, Zhengzhou, Henan 451191, China

E-mail: baojian134597@yeah.net

Keywords: data mining, library, book borrowing, radial basis function neural network, prediction

Received: February 3, 2021

*In the library, the prediction and estimation of book borrowing plays an important role in library work. Based on the data mining method, this paper analyzed the prediction and estimation of book borrowing. Firstly, the radial basis function neural network (RBFNN) was analyzed. Then, the improved ant colony algorithm (IACO) was used to obtain the optimal parameters of RBFNN, and then the IACO-RBFNN model was established to realize the prediction and estimation of book borrowing. The results showed that the improved model had advantages in training time, iteration times, and error compared with BPNN and RBFNN. The results of book prediction and estimation showed that the results obtained by the IACO-RBFNN model were closer to the actual book borrowing situation, with smaller error and higher precision (97.09%), and its precision was 11.18% and 4.74% higher than BPNN and RBFNN respectively. The training time and testing time of the IACO-RBFNN model were 5.12 s and 1.03 s, respectively, which were significantly shorter than the other two methods. The results show that the IACO-RBFNN model has a good performance in the prediction and estimation of book borrowing and can be further promoted and applied in practice.*

*Povzetek: Opisana je metoda strojnega učenja za napovedovanje izposoje knjig v knjižnici.*

## 1 Introduction

The library is an important facility in a school. It can meet the needs of teachers and students in teaching and scientific research by collecting and sorting books. With the expansion of the school scale, the amount of books borrowed in the library is also growing. In the management of the library, the borrowing amount can reflect the work quality of the library to a certain extent and has a reference value for the purchase of new book resources. Therefore, it is of great significance to predict, estimate, and analyze the borrowing amount [1]. Data mining refers to the process of finding hidden and useful information from massive data, which has been widely used in data prediction and estimation [2]. Shan et al. [3] studied the on-line prediction of tool wear, designed a method based on least squares support vector machine regression, and found through experiments that the method had better accuracy than a neural network. Qazi et al. [4] analyzed the role of artificial neural networks in predicting solar radiation. Through the analysis of 24 literature, they found that the prediction error of the artificial neural network was smaller than 20% and it could process a variety of input meteorological parameters. Zhang et al. [5] combined the long short-term memory method with the recurrent neural network to predict the remaining life of lithium-ion batteries. Through experiments and comparison, they found that the method could predict the remaining life of the lithium-ion batteries effectively. Manek et al. [6] used the back propagation neural network (BPNN), generalized

regression neural network (GRNN), and radial basis function neural network (RBFNN) to predict the rainfall in Thanjavur district of southern province Tamil Nadu, India, and found that the RBFNN could get the best prediction results. Ramos et al. [7] predicted delayed cerebral ischemia (DCI) in patients with aneurysmal subarachnoid hemorrhage, combined logistic regression model, machine learning model, and automatic encoder, trained and tested the model with 317 cases of data, and found that the method could effectively improve the prediction of DCI in patients. Souri et al. [8] studied the fault prediction of the Internet of things and proposed a model combining multi-layer perceptron and particle swarm optimization algorithm. The experiment showed that the method had short operation time and small memory consumption. Aiming at the problem of urban traffic flow prediction, Hu et al. [9] established GSTAR-SVM model with wavelet transform and predicted the short-term traffic flow. Through experiments, they found that the model had high prediction accuracy. Iqbal et al. [10] evaluated the performance of seven machine learning methods in predicting dengue outbreak and found through an experiment that the LogitBoost integration model had the highest classification accuracy (92%), a sensitivity of 90%, and a specificity of 94%. At present, the application of methods such as data mining and machine learning in library management is seldom, and artificial method is highly dependent, which is not conducive to the scientific management of a large number of books. Therefore, based on RBFNN, this study applied RBFNN to the prediction and estimation of book borrowing and optimized it with

the ant colony optimization (ACO) algorithm to improve the accuracy of prediction and estimation. This work aims to guide book purchase and management of libraries.

## 2 Book borrowing prediction and estimation model

### 2.1 RBF neural network

The problem of book borrowing prediction and estimation is affected by many factors and has nonlinear characteristics. However, the models used in this problem, such as the regression analysis model [11] and grey model [12], are all linear estimation models, which has poor estimation accuracy. A neural network is a kind of data mining, which is a simulation of a biological neural network. A neural network is a kind of nonlinear estimation model with excellent nonlinear approximation ability [13]. BPNN [14] and RBFNN [15] have been widely used in prediction and estimation. Compared with BPNN, RBFNN has more advantages in operation speed and structure and has been successfully applied in fields such as human face recognition [16] and defect detection [17]. Therefore, this study used RBFNN to establish the prediction and estimation model of book borrowing.

RBFNN is a three-layer forward network. It is assumed that the input layer of RBFNN has  $n$  nodes,  $X = (x_1, x_2, \dots, x_n)$ , its output layer has  $m$  nodes,  $Y = (y_1, y_2, \dots, y_m)$ , its hidden layer has  $h$  nodes, then the output of RBFNN can be written as:

$$y_i = f_i(x_i) = \sum_{k=1}^N w_{ik} \phi_k(\|x - c_k\|_2), i = 1, 2, \dots, m,$$

where  $w_{ik}$  refers to the weight between the hidden layer and output layer,  $\phi_k$  is the activation function, and  $c_k$  is to the center vector of the basis function. In RBFNN, the most commonly used activation function is Gaussian function. Compared with other functions, the Gaussian function is simpler and radial symmetric and has better smoothness. The formula of the Gaussian function is:

$$\phi(x) = \exp\left(-\frac{x^2}{2\sigma^2}\right),$$

where  $\sigma^2$  is a variance. In this case, the output of RBFNN can be written as:

$$y_i = f_i(x_i) = \sum_{k=1}^N w_{ik} \exp\left[-\left(\frac{\|x - c_k\|_2^2}{2\sigma^2}\right)\right], i = 1, 2, \dots, m.$$

To sum up, it can be found that parameters  $w_{ik}$ ,  $c_k$ , and  $\sigma$  have a great influence on the performance of RBFNN, which is also the key and difficult point to establish the RBFNN model. In order to find the optimal parameters, this study selected the ACO algorithm.

### 2.2 Ant colony algorithm

ACO algorithm is a heuristic algorithm based on simulated ant colony behaviors [18]. In the process of ants' foraging, pheromones will be released. In the process of searching for paths, ants will find the path with high pheromone concentration and release pheromone at the

same time, which will make the pheromone concentration on the path higher and higher, and all ants will gather on one path finally. For ant  $k$ , the probability of ant  $k$  from city  $i$  to city  $j$  at time  $t$  can be written as:

$$\rho_{ij}^k(t) = \begin{cases} \frac{\delta_{ij}^{\alpha(t)} \eta_{ij}^{\beta}}{\sum_{j \in N_j^k} \delta_{ij}^{\alpha(t)} \eta_{ij}^{\beta}}, j \in allowed_k \\ 0, otherwise \end{cases}$$

The update process of pheromone can be written as:

$$\delta_{ij}(t + n) = p\delta_{ij}(t) + \sum_{k=1}^m \Delta\delta_{ij}^k,$$

$$\Delta\delta_{ij}^k = \begin{cases} L_k^{-1}, j \in allowed \\ 0, otherwise \end{cases}$$

In the above formula, the parameters involved and their meanings are shown in Table 1.

Parameter	Meaning
$\delta_{ij}$	Pheromone concentration
$\eta_{ij}$	Heuristic factor
$\alpha$	The importance of pheromone
$\beta$	The importance of heuristic factor
$N_j^k$	A city without passing by
$allowed_k$	Feasible solution set
$p$	Pheromone volatilization factor
$L_k$	The length of a path that ant $k$ passes

Table 1: Parameter table.

In the ACO algorithm, the value of volatilization factor  $p$  is between 0 and 1. When the value of  $p$  is too large, it may affect the global search ability of the algorithm. Therefore, this study used an adaptive method to improve the ACO algorithm. The initial value of  $p$  is set as 0.9, and then it changes followed the following formula:

$$p_t = \begin{cases} 0.9p(t - 1), 0.9p(t - 1) \geq p_{low} \\ p_{low}, p_{low} \end{cases}$$

where  $p_{low}$  is the minimum value of  $p$ .

### 2.3 Improved ACO-RBFNN model

The parameters of RBF were optimized using the improved ACO (IACO) algorithm. It is assumed that there are  $m$  parameters including  $w_{ik}$ ,  $c_k$ , and  $\sigma$ , and they are randomly sorted, which is set as  $R_i, i \in [1, m]$ .  $R_i$

is taken as the food source, and the optimal parameters are searched according to the IACO algorithm. When all the ants concentrated on the same route, the parameters obtained at that moment were optimal for RBF. The flow

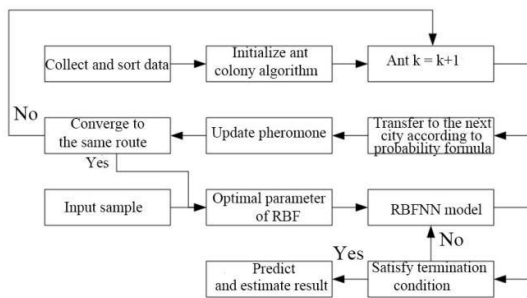


Figure 1: The IACO-RBFNN model.

chart of the IACO-RBFNN model designed for book borrowing prediction and estimation is shown in Figure 1.

As shown in Figure 1, the optimal parameters of RBFNN are obtained using the IACO algorithm after the collected data are processed, those parameters are used for establishing the RBFNN model, and the model is trained by inputting training samples until the model iterates out the most accurate result. The obtained result is the prediction and estimation result of book borrowing.

### 3 Experimental analysis

#### 3.1 Experimental data

Taking the Graphic Center of Henan Mechanical and Electrical Vocation College as an example, the book borrowing data from January 2018 to June 2020 (24 months) were collected through the library information system. During the training of the IACO-RBFNN model, the data from January 2018 to December 2019 was used as training samples, the number of books borrowed in one month as one sample. The data of the fourth month were predicted based on the data of the first three months, for example, estimating the data of April 2018 based on the data of January 2018 ~ March 2018, i.e., the number of books borrowed in January, February, and March 2018 were taken as the input of the RBFNN model, and the number of books borrowed April 2018 was taken as the

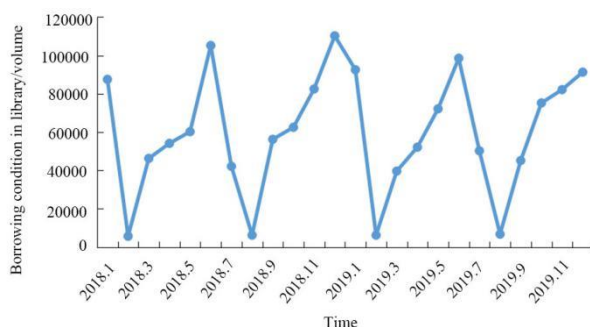


Figure 2: Data of the training sample.

output of the RBFNN model. The training sample is shown in Figure 2.

It was seen from Figure 2 that there is a law in the borrowing of books. In January, June, and December of each year, the number of books borrowed is relatively large, while the number in February and August is small. The above phenomenon may be related to the particularity of the school. In the months of the final examination, the borrowing demand for books is great, while the borrowing demand significantly decreases during the winter and summer vacation.

In order to speed up the operation of the formula, it is necessary to standardize the collected data using the following formula:

$$x' = \frac{x - x_{min}}{x_{min} - x_{max}}$$

where  $x'$  refers to the normalized data,  $x$  is the original data, and  $x_{max}$  and  $x_{min}$  are the maximum and minimum values of the original data.

#### 3.2 Experimental results

Firstly, a nonlinear function,  $y = f(x)$ ,  $x \in [-1, 1]$ , was used to verify the performance of the IACO-RBFNN model, and it was compared with the traditional BPNN and traditional RBFNN models. The target accuracy was 0.01, and the maximum number of iterations was 500. The

	BPNN	RBFN N	IACO- RBFNN
Training time/s	36.78	10.16	2.34
Number of iterations	245	148	34
Error	1.26	0.53	0.26

Table 2: Comparison of model performance.

performance of the three methods is shown in Table 2.

It was seen from Table 2 that the IACO-RBFNN model had obvious advantages in performance. Firstly, in terms of training time, the BPNN model took 36.78 s in training, the RBFNN model took 10.16 s, and the IACO-RBFNN model only took 2.34 s, which was significantly shorter than the other two models; secondly, in terms of the number of iterations, the BPNN model needed 245 times of iterations to get the optimal value, the RBFNN model needed 148 times, and the IACO-RBFNN model only needed 34 times; finally, from the perspective of error, BPNN model > RBFNN model > IACO-RBFNN model. In a comprehensive view, the model designed in this paper had the best performance.

In the book borrowing prediction and estimation from January 2020 to June 2020, the results of the three models are shown in Figure 3.

It was seen from Figure 3 that there was a gap between the estimated results of the BPNN and RBFNN models and the actual borrowing situation. The error of the

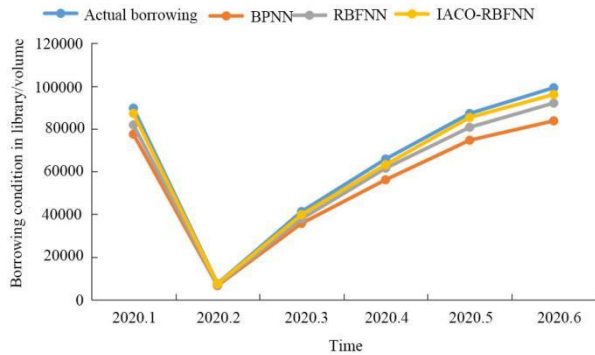


Figure 3: Book borrowing prediction results of three models.

RBFNN model was larger than the BPNN model, which indicated that the RBFNN model had a better performance. Compared with the RBFNN model, the estimated result of the IACO-RBFNN model was closer to the actual borrowing result, which showed that the RBFNN model had a significantly improved performance after improvement by the IACO algorithm, and it had better accuracy in the prediction and estimation of book borrowing. In order to further verify the effectiveness of the model designed in this study, the error and precision of the three models were calculated, and the results are shown in Table 3.

It was seen from Table 3 that the error of the BPNN model was the largest, while that of the IACO-RBFNN model was the smallest. In the prediction and estimation, the average error of the BPNN, RBFNN, and IACO-RBFNN models was 9404 books, 4967 books, and 1955 books, respectively, and the average error of the IACO-RBFNN model was 79.21% less than that of the BPNN model and 60.64% less than that of the RBFNN model. The average precision of the BPNN and RBFNN models was 85.91% and 92.35%, respectively, while the average precision of the IACO-RBFNN model was 97.09%, which

	BPNN model		RBFNN model		IACO-RBFNN model	
	Error	Precision	Error	Precision	Error	Precision
January 2020	12175	86.42%	7836	91.26%	2457	97.26%
February 2020	962	87.26%	577	92.35%	139	98.16%
March 2020	5553	86.54%	3437	91.67%	1452	96.48%
April 2020	9702	85.26%	4305	93.46%	2646	95.98%
May 2020	12516	85.64%	6467	92.58%	1874	97.85%
June 2020	15518	84.36%	7183	92.76%	3165	96.81%
Average value	9404	85.91%	4967	92.35%	1955	97.09%

Table 3: Comparison results of error and precision.

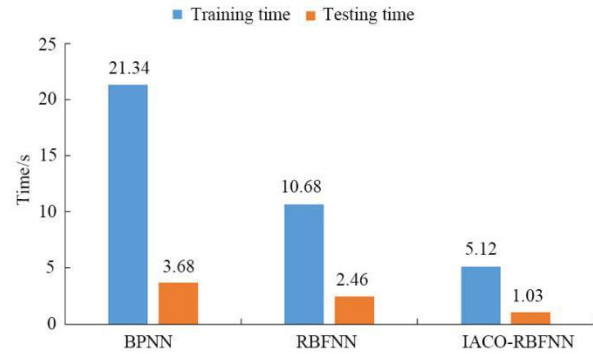


Figure 4: Comparison of operation time.

was 11.18 higher than the BPNN model and 4.74% higher than the RBFNN model. Thus, it was concluded that the IACO-RBFNN model was the most effective in the prediction and estimation of book borrowing.

Finally, the operation time of the model was compared, as shown in Figure 4.

It was seen from Figure 4 that the operation time of the BPNN model was the longest, followed by the RBFNN and IACO-RBFNN models. The training time and testing time of the BPNN model were 21.34 s and 3.68 s, respectively; the operation time of the RBFNN model significantly reduced. The training time of the IACO-RBFNN model was 5.12 s, which improved 52.06% compared to the RBFNN model, and the testing time of the model was 1.03 s, which improved 58.13% compared to the RBFNN model. It was found that the algorithm improved by the IACO algorithm had improved precision and significantly shortened operation time and showed a better performance in the prediction and estimation of book borrowing.

## 4 Discussion

The development of information technology has brought new changes to many fields. Many industries have established information systems to realize information management, so does the library [19]. In the process of library informatization, a large amount of information has been accumulated, but most digital libraries cannot effectively develop and utilize these data and make the information accumulate, which brings great difficulties to the resource management and data processing of the library. In order to develop the library better, a method is urgently needed to realize the analysis and utilization of these data, and the emergence of data mining solves this problem [20].

This study mainly analyzed RBFNN. For the parameter selection of RBFNN, many algorithms have been applied, such as the gravity search algorithm [21], genetic algorithm [22], grey wolf optimization algorithm [23], etc. This paper selected the ACO algorithm, improved the ACO algorithm to optimize the parameters of RBFNN, carried out experiments with the actual book lending data, and compared the IACO-RBFNN model with the BPNN and RBFNN models. The results suggested that the IACO-RBFNN model needed fewer times of iterations, shorter training time, and smaller error compared to the BPNN and RBFNN models, indicating

that the IACO-RBFNN model had more obvious advantages in performance. In the prediction and estimation of book borrowing, the prediction result of the RBFNN model was closer to the actual situation than that of the BPNN model, suggesting that the performance of the RBFNN model was better than that of BPNN. Then, the comparison between the IACO-RBFNN model and the RBFNN model found that the prediction result of the IACO-RBFNN model was closer to the actual situation. It was seen from Table 3 that the IACO-RBFNN model had smaller prediction error and higher precision, and the average precision of the method was 97.09%, which was 11.18% higher than that of the BPNN model and 4.74% higher than that of the RBFNN model. In the comparison of the operation time, the IACO-RBFNN model was significantly shorter, i.e., it could obtain results with high precision in a short time, which showed that the IACO-RBFNN model had better usability in the prediction and estimation of book borrowing.

Although some useful achievements have been made in this article, there are some deficiencies that need to be solved in future work:

- (1) comparing the performance of more data mining methods;
- (2) further optimizing the precision of the RBFNN model;
- (3) studying more applications of data mining methods in libraries.

## 5 Conclusion

This study designed the IACO-RBFNN model for the prediction and estimation of book borrowing in the library. Taking the data in the Graphic Center of Henan Mechanical and Electrical Vocation College as an example, the experiment was carried out. The comparison with the BPNN and RBFNN models found that:

- (1) the IACO-RBFNN model needed shorter training time and fewer times of iterations and had smaller error;
- (2) the predicted result of the IACO-RBFNN model was closer to the actual book borrowing situation;
- (3) the average error and average precision of the IACO-RBFNN model was 1955 books and 97.09%, which was better than the other two models;
- (4) the training time and testing time of the IACO-RBFNN model were only 5.12 s and 1.03 s, respectively.

It was found from the results that the IACO-RBFNN model had a good performance in the prediction and estimation of book borrowing and could be applied in the actual library work to give guidance for the library work.

## References

- [1] Xu SX (2020). Association rule model of on-demand lending recommendation for university library. *Informatica*, 44, pp. 395-399. <https://doi.org/10.31449/inf.v44i3.3295>.
- [2] Javeed S, Mohamed S (2020). Novel Feature Reduction (NFR) Model With Machine Learning and Data Mining Algorithms for Effective Disease Risk Prediction. *IEEE Access*, 8, pp. 184087-184108. <https://doi.org/10.1109/ACCESS.2020.3028714>.
- [3] Guan S, Yan LH, Peng C (2015). Application of regression algorithm of LS-SVM in tool wear prediction. *China Mechanical Engineering*, 26, pp. :217-222. <https://doi.org/10.3969/j.issn.1004-132X.2015.02.016>.
- [4] Qazi A, Fayaz H, Wadi A, Raj R, Rahim N, Khan W (2015). The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review. *Journal of Cleaner Production*, 104, pp. 1-12. <https://doi.org/10.1016/j.jclepro.2015.04.041>.
- [5] Zhang Y, Xiong R, He H, Pecht MG (2018). Long Short-Term Memory Recurrent Neural Network for Remaining Useful Life Prediction of Lithium-Ion Batteries. *IEEE Transactions on Vehicular Technology*, 67, pp. 5695-5705. <https://doi.org/10.1109/TVT.2018.2805189>.
- [6] Manek AH, Singh PK (2016). Comparative study of neural network architectures for rainfall prediction. pp. 171-174. <https://doi.org/10.1109/TIAR.2016.7801233>.
- [7] Ramos L A, Steen W E V D, Barros R S, Majoie CBLM, van den Berg R, Verbaan D, Vandertop WP, Zijlstra IJAJ, Zwinderman AH, Strijkers GJ, Olabarriaga SD, Marquering HA (2019). Machine learning improves prediction of delayed cerebral ischemia in patients with subarachnoid hemorrhage. *Journal of Neurointerventional Surgery*, 11, pp. 497-502. <https://doi.org/10.1136/neurintsurg-2018-014258>.
- [8] Souri A, Mohammed A S, Potrus M Y, Malik MH, Safara F, Hosseinzadeh M (2020). Formal Verification of a Hybrid Machine Learning-Based Fault Prediction Model in Internet of Things Applications. *IEEE Access*, 8, pp. 23863-23874. <https://doi.org/10.1109/ACCESS.2020.2967629>.
- [9] Hu J, Wang S, Mao J (2019). Research on GSTAR-SVM Traffic Prediction Model Based on Wavelet Transform. *Journal of Physics: Conference Series*, 1345, pp. 032009 (7pp). <https://doi.org/10.1088/1742-6596/1345/3/032009>
- [10] Iqbal N, Islam M (2019). Machine learning for dengue outbreak prediction: A performance evaluation of different prominent classifiers. *Informatica*, 43, pp. 363-371. <https://doi.org/10.31449/inf.v43i3.1548>.
- [11] Wang CY, Wang WS (2015). Regression Analysis When Covariates Are Regression Parameters of a Random Effects Model for Observed Longitudinal Measurements. *Biometrics*, 56, pp. 487-495. <https://doi.org/10.1111/j.0006-341X.2000.00487.x>.
- [12] Fu M, Wang W, Le Z, Khorram MS (2015). Prediction of particular matter concentrations by developed feed-forward neural network with rolling mechanism and gray model. *Neural Computing & Applications*, 26, pp. 1789-1797. <https://doi.org/10.1007/s00521-015-1853-8>.
- [13] Boufadene M, Belkheiri M, Rabhi A, Hajjaji AE (2019). Vehicle longitudinal force estimation using

- adaptive neural network nonlinear observer. *International Journal of Vehicle Design*, 79, 205-.  
<https://doi.org/10.1504/IJVD.2019.103593>.
- [14] Lyu J, Zhang J (2018). BP Neural Network Prediction Model for Suicide Attempt among Chinese Rural Residents. *Journal of Affective Disorders*, 246, pp. 465-473.
- [15] Xiong T, Bao Y, Hu Z, Chiong R (2015). Forecasting interval time series using a fully complex-valued RBF neural network with DPSO and PSO algorithms. *Information Sciences*, 305, pp. 77-92.  
<https://doi.org/10.1016/j.ins.2015.01.029>.
- [16] Dey A, Ghosh M (2019). A Novel Approach to Fuzzy-Based Facial Feature Extraction and Face Recognition. *Informatica*, 43, pp. 535-543.  
<https://doi.org/10.31449/inf.v43i4.2117>.
- [17] Jiang HN (2018). Defect features recognition in 3D industrial CT images. *Informatica*, 42, pp. 477-482.  
<https://doi.org/10.31449/inf.v42i3.2454>.
- [18] Mandloi M, Bhatia V (2015). Congestion control based ant colony optimization algorithm for large MIMO detection. *Expert Systems with Applications*, 42, pp. 3662-3669.  
<https://doi.org/10.1016/j.eswa.2014.12.035>.
- [19] Day A (2018). Research Information Management: How the Library Can Contribute to the Campus Conversation. *New Review of Academic Librarianship*, 24, pp. 23-34.  
<https://doi.org/10.1080/13614533.2017.1333014>.
- [20] Wang W, Meng L, Wu L, Zhang J. (2020). Research and Application of Data Mining Technology in Library Office Information Construction. *Journal of Physics: Conference Series*, 1550, pp. 032001 (6pp).  
<https://doi.org/10.1088/1742-6596/1550/3/032001>.
- [21] Assareh E, Biglari M (2015). A novel approach to capture the maximum power from variable speed wind turbines using PI controller, RBF neural network and GSA evolutionary algorithm. *Renewable & Sustainable Energy Reviews*, 51, pp. 1023-1037.  
<https://doi.org/10.1016/j.rser.2015.07.034>.
- [22] Jia W, Zhao D, Ding L (2016). An optimized RBF neural network algorithm based on partial least squares and genetic algorithm for classification of small sample. *Applied Soft Computing*, pp. 373-384.  
<https://doi.org/10.1016/j.asoc.2016.07.037>.
- [23] Shang S, He K N, Wang Z B, Yang T, Liu M, Li X. (2020). Sea Clutter Suppression Method of HFSWR Based on RBF Neural Network Model Optimized by Improved GWO Algorithm. *Computational Intelligence and Neuroscience*, 2020, pp. 1-10.  
<https://doi.org/10.1155/2020/8842390>.