

Research on Financial Early-warning based on GIHS Improved BP_AdaBoost Algorithm

Gui-fang Shen, Yi-Wen Zhang

School of Information Engineering, Anhui Xinhua University, Hefei China ,230088

Abstract. To improve the accuracy of the financial early warning of the company, aiming at defects of slow learning speed, trapped in local solution and inaccurate operating result of the traditional BP neural network with random initial weights and thresholds, a parallel ensemble learning algorithm based on improved harmony search algorithm using good point set (GIHS) optimize the BP_Adaboost is proposed. Firstly, the good-point set is used to construct a more high quality initial harmony library, and it adjusts the parameters dynamically during the search process and generates several solutions in each iteration so as to make full use of information of harmony memory to improve the global search ability and convergence speed of algorithm. Secondly, ten financial indicators are chosen as the inputs of BP neural network value, and GIHS algorithm and BP neural network are combined to construct the parallel ensemble learning algorithm to optimize BP neural network initial weights value and output threshold value. Finally, many of these weak classifier is composed as strong classifier through the AdaBoost algorithm. The improved algorithm is validated in the company's financial early warning. Simulation results show that the performance of GIHS algorithm is better than the basic HS and IHS algorithm, and the GIHS-BP_AdaBoost classifier has higher classification and prediction accuracy.

1. Introduction

With the development of the market economy in our country, enterprises need to give a comprehensive and timely warning of their financial status before they can calmly deal with various crises and challenges. Financial Forecast is a financial analysis system that early warning the enterprise from the financial point of view, early detection of the crisis facing the enterprise by analyzing the financial statements and related data indicators of the enterprise, analyzing the causes and making corresponding preventive measures.[1] Financial early warning research methods are divided into qualitative analysis of early warning, quantitative analysis and early warning and data mining analysis and early warning. In recent years, the mainstream data mining analysis method uses mathematical and computer software to build the prediction model. Starting from the specific data analysis, it has higher prediction accuracy than the qualitative and quantitative warning methods. Among them, the neural network has become one of the mainstream analysis models. Reference [2] first improved the harmony search algorithm by combining the differential evolution algorithm, and then used the improved algorithm to optimize the BP neural network parameters to build the HSDM-BP model. Experiments based on financial data show that the model predicts the financial crisis of enterprises is more accurate than other methods. In the literature [3], firstly, the financial index

of the company was reduced based on the neighborhood rough set, and the weights and thresholds of the BP neural network were optimized by the particle swarm optimization algorithm. The experiment proved that the model is effective and feasible for the financial early warning of enterprises.

Back Propagation Neural Networks (BPNN) Due to its relatively simple network topology and good nonlinear mapping capability, BPNN has outstanding advantages in solving some nonlinear problems. However, the random initial weights and thresholds of BP algorithm and learning methods based on gradient descent result in slow convergence of BP algorithm and often suffer from local minima. At present, many scholars use genetic algorithms [4], artificial bee colony algorithm [5], glowworm swarm optimization algorithm [6] and other intelligent algorithms to optimize the serial combination of BP network parameters, intelligent algorithm optimization and BP network gradient descent Improve performance. The optimized network has a certain improvement in the calculation accuracy, but still belongs to the weak classifier. Literature [7] exploited the global search performance of genetic algorithm to optimize BP neural network initial connection weights and output thresholds, and then integrated several weak classifiers of BP neural network to construct AdaBoost strong classifier and applied it to traffic event prediction Experimental results show that the integrated algorithm can improve the performance of the original BP weak classifier.

The Harmony Search Algorithm (HSA) [8] stems from the imitation of the songwriting process. Its simple structure and strong global search capability have been applied to multidimensional multi-pole optimization, pipeline optimization design, water resource network design, scheduling and 0-1 knapsack [9-12]. However, the search mechanism of HSA determines its randomness and slow convergence speed, and has the disadvantage that it is easy to fall into the local optimum and has unstable convergence in the late stage. The researchers made some improvements to the algorithm [13-15].

Therefore, this paper presents the corporate financial alert model of GIHS-BP_AdaBoost. Firstly, the harmony search algorithm is improved based on the good point set, the construction of a uniform initial sound library and the improvement of the key parameters, and then the improved harmony search algorithm (GIHS) based on the good point set is combined with the BP neural network to construct parallel Learning algorithms to optimize the network structure and connection weights of neural networks and obtain several classifiers. Finally, several optimized BP neural network weak classifiers are built into strong classifiers by AdaBoost algorithm and applied to the company's financial early warning.

2. Improved Harmonic Search Algorithm

In order to improve the optimization performance of the basic HSA, an improved harmonic search algorithm based on good point set (GIHS) is proposed.

2.1 Optimal Initialization of Harmony Memory based on good point set

In order to solve the problem that the harmony in the initial harmony library is unevenly distributed in the search space, the initial harmony memory is constructed based on the good point set [16]. According to the research results in [17] and [18], it can be seen that the distribution of population generated by the good point set technique is more uniform than the random initialized population distribution, and the population diversity is more abundant.

Definition 1 Suppose there is a cube V_S in the S -dimensional Euclidean space, and the existence form is as $P_n(k) = \{(r_1^{(n)} \times k), (r_2^{(n)} \times k), \dots, (r_s^{(n)} \times k), 1 \leq k \leq n\}$, and $r \in V_S$. If its deviation $\Phi(n) = C(r, \varepsilon)n^{-1+\varepsilon}$ holds, where ε is an arbitrary positive integer and $C(r, \varepsilon)$ is a constant related only to r and ε , then r is a good point, $P_n(k)$ is a good point set.

In general, take $r_k = \{2\cos(2\pi k/p), 1 \leq k \leq s\}$, p is the smallest prime number that satisfies $(p-1)/2$, or

$$r_k = \{\exp(k)\}, 1 \leq k \leq s. \quad (1)$$

Where $\{a\}$ represents the fraction of a .

2.2 Generate new harmonic parameters dynamically adjusted

In order to improve the algorithm convergence speed and search accuracy, the pitch adjustment probability

value P_{AR} is dynamically adjusted from small to large, and the harmony fine adjustment amplitude value B_W is dynamically adjusted from big to small. The learning and sound library probability H_{MCR} is dynamically adjusted from small to large.

The expressions for dynamic H_{MCR} , P_{AR} and B_W are given in [14] and [15].

2.3 Adjustment of generating new harmonies in each generation

The basic HSA generates only one new harmony per iteration for updating the harmony library, and the update is too small. In order to make full use of the historical information of the harmony library, GIHS generates N new harmonies for each iteration, puts them in the original harmony library together with H_{ms} harmonies, and finds out the H_{ms} optimal solutions as the new harmony Memory bank.

3. GIHS-BP_AdaBoost Classification algorithm

3.1 GIHS-BPParallel integrated learning algorithm

The basic idea of GIHS-BP parallel integrated learning algorithm is: to determine the structure of BP neural network based on input and output parameters to determine the encoding length of each individual harmony. Each individual in the population contains BP's ownership and thresholds, calculates individual fitness values, and searches for individuals with best fit fitness values through improved harmony search. The individuals obtained by GIHS optimization are assigned initial weights and thresholds of BP network, and then optimized by BP network model to form a learning algorithm integrated with GIHS algorithm and gradient descent method of BP neural network.

The basic steps are described as follows:

Step1: Coding, initializing the harmony memory. For n input nodes, l output nodes and m hidden layer nodes of BP neural network, there are $n*m+m*l$ weights and $m+l$ thresholds, each of which includes BP threshold and weight all parameters, can represent a BP network structure. And through the formula (1) the use of good points set preferred merit memory structure initial memory.

Step2: Initializing the Parameters. The parameters include the size of the harmony memory bank H_{ms} , the number of new solutions N in each batch, the upper and lower limits of pitch adjustment probability P_{ARmax} and P_{ARmin} , the upper and lower bounds H_{MCRmax} and H_{MCRmin} of the learning harmony sound library, the maximum number of iterations T , the adjustment bandwidth B_{Wmin} and B_{Wmax} , The number of decision variables N_{VAR} (equal to the sum of all BP system parameters).

Step3: fitness function calculation. Fitness is the main index to describe the quality of the population. In this paper, root mean square error is chosen as fitness

function, as shown in formula (2):

$$f = \left\{ \frac{1}{S-1} \sum_{i=1}^S (t_p - y_p)^2 \right\}^{1/2} \quad (2)$$

In the formula, S is the prediction sample number, t_p is the expected output of the p th sample, y_p is the actual output of the p th sample.

Step4: New harmonies are generated and N new harmonies are obtained for each iteration.

Step5: Update harmony memory. The fitness function values of each new harmony are calculated, and H_{ms} good ones are selected to constitute a new harmony library from the $N+H_{ms}$ newly generated new solutions and the original harmony memory bank.

Step6: judge the algorithm termination condition. If the current number of iterations t is equal to the maximum number of iterations T , then terminate the operation; otherwise repeat Step4, Step5.

Step7: decompose the harmony individuals obtained by GIHS optimization into the network parameters of the BP neural network, and use the trained BP neural network to predict, and obtain the optimal results of the company's financial warning classification and prediction.

3.2 GIHS-BP_AdaBoost algorithm

The idea of AdaBoost algorithm is to integrate the output of multiple weak classifiers to produce the desired result with less error. In order to enhance the classification effect, the results obtained by the parallel learning of GIHS-BP are taken as the new weak classifier, and then the GIHS-BP_AdaBoost strong classifier is constructed by AdaBoost algorithm.

First of all, select the L group data from the sample space as the training samples, initialize the weights of the training samples, according to the following formula:

$$D_i(r) = 1/L (r \in 1, 2, \dots, L) \quad (3)$$

Then the BP neural network optimized by GIHS is trained by the training sample data. When training the k -th ($k \in 1, 2, \dots, M$) BP neural network weak classifier, the prediction error of the group data is obtained as follows:

$$e_k = \sum [g_k(r) - y_r] (r \in 1, 2, \dots, L) \quad (4)$$

Where: $g_k(r)$ is the expected output value of the network, y_r is the actual output of the network.

Then the weight of the k -th ($k \in 1, 2, \dots, M$) BP neural network weak classifier is calculated according to the prediction error of n training samples and e_k as follows:

$$a_k = \frac{1}{2} \ln \left(\frac{1 - e_k}{e_k} \right) \quad (5)$$

Then, according to the weight a_k , adjust the weight of the next round of training samples, the formula is:

$$D_{k+1}(r) = \frac{D_k(r)}{B_k} \exp[-a_k g_k(r) y_i] (r \in 1, 2, \dots, L) \quad (6)$$

Where: B_k is the normalization factor, $D_k(i)$ is the weight after $k-1$ training and adjusted.

Finally, a strong classification function is established. After T times of training, a strong classification function is obtained by combining the weak classification

functions of T groups. The calculation formula is:

$$h(x) = \sum_{k=1}^T a_k f_k(x) \quad (7)$$

Where: $f_k(x)$ is the k -th weak classification function; x is the test data.

4. Company Financial Early-warning Model Based on GIHS-BP_AdaBoost

4.1 Establishment of Index System for Financial Early-warning of the Company

To build an efficient enterprise financial early warning model, we must choose to establish a reasonable indicator system firstly. The financial warning indicator system mainly includes indicators such as corporate profitability, solvency, development capability and operational capability. The main indicators to evaluate the profitability of enterprises are: sales interest rate, profit margin of main business, net margin of assets, main profit proportion and net cash flow of operating activities per share. The indicators for evaluating and analyzing the short-term solvency of enterprises are: liquidity ratio and quick ratio. The indicators for evaluating the long-term solvency of enterprises include such indicators as asset-liability ratio, working capital proportion, long-term debt and working capital ratio. Indicators that reflect the development of enterprises mainly include indicators of capital accumulation rate, total assets growth rate, net profit growth rate and growth rate of main business revenue. If all indicators are included in the financial early warning evaluation, the model will be too complicated. And the correlation between the indicators is strong, so before the model needs to be screened.

Based on the research of financial crisis early warning index selection in [1-3], this paper further finds the index with large eigenvalue as the final evaluation index of the company's crisis early warning method as follows: Component cost profit margin, Assets operational capacity, Total assets growth rate, The total assets of the company, Current ratio, Operating cash flow, Audit opinion type, Earnings per share, Inventory turnover, Assets and liabilities

4.2 Simulation

BP neural network determines the number of network input and output nodes according to the input and output dimension of training samples. The model chooses 10 indicators, that means the input data dimension is 10, the output node is 1, and the number of hidden layer nodes is $2 * 10 + 1$ equals 21. Therefore, the network structure constructed is 10-21-1.

4.2.1 Data sampling and processing.

This case contains a total of 1350 group company financial status data, The input for each set of data is 10

dimensions, representing the above 10 indicators, The output is 1-dimensional, representing the company's financial position. 1 means good financial position, -1

indicates a financial problem. Randomly select 1000 sets of data as training data, 50 groups for the forecast data. Part of experimental data is in Table 1.

Table 1 Some sample data of BoZhou area

sample	index										grade
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	Y
S ₁	2	4000	8	1	2	51000	36	1	3	23	1
S ₂	3	5200	10	1	2	150000	36	1	3	23	1
S ₃	2	5000	6	5	2	83000	36	1	3	38	-1
S ₄	2	18000	10	1	2	203000	36	1	3	42	1
S ₅	1	10000	10	5	2	230000	36	1	3	27	-1
S ₆	3	4000	10	5	2	7488	36	1	3	29	-1
S ₇	3	5200	10	1	2	150000	36	1	3	23	1
S ₈	2	5000	10	1	2	99000	60	1	3	23	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
S ₂₀₀	2	3000	10	5	2	76000	36	1	3	33	-1

The GIHS-BP_Adaboost strong classifier forecasting model proposed in this paper is applied to the company's financial early-warning and is compared with weak classifier BP neural network prediction model, GIHS optimized BP neural network prediction model, basic harmony algorithm optimization BP (HS-BP_Adaboost) and The improved BP prediction model (IHS-BP_Adaboost) proposed by [14] and acoustic search algorithm is compared to verify the feasibility and effectiveness of the proposed algorithm.

4.2.2 Experimental parameters set.

The HS, IHS and GIHS algorithm parameters are set as follows: $H_{ms} = 30$, $N = 5$, $T = 5000$, and other parameters are set as in Table 2. BP neural network has a structure of 10-21-1, so there are $10 * 21 + 21 * 1 = 231$ weights and $10 + 1 = 11$ thresholds. Therefore, the number of decision variables of the harmony search and the improved harmony search algorithm is $N_{VAR} = 231 + 11 = 242$. The parameters of BP neural network are generally between -0.5 and 0.5, so the feasible domain of initial harmony $\in [-0.5, 0.5]$. The number of weak class predictors is set to 20.

Table2 Parameter settings for the compared algorithms

Algorithms	Parameters and settings
HS	$H_{MCR}=0.9$, $P_{AR}=0.35$, $B_w=0.01$
IHS	$H_{MCR} =0.9$, $P_{ARmax} =0.9$, $P_{ARmin} =0.4$, $B_{Wmin}=0.001$, $B_{Wmax}=1.0$
GIHS	$P_{ARmax} =0.9$, $P_{ARmin} =0.4$, $H_{MCRmax} =0.99$, $H_{MCRmin} =0.6$, $B_{Wmin} =0.001$, $B_{Wmax} =1.0$,

4.2.3 Experimental results and analysis.

The financial early warning of the company is classified and predicted by adopting the weak classifier (BP, HS-BP, IHS-BP, GIHS-BP) and the strong classifier (BP_Adaboost, HS-BP_Adaboost, IHS-BP_Adaboost,

GIHS-BP_Adaboost) Output classification result and forecast result.

In order to compare the optimization performance of the improved harmony search algorithm, the iterative convergence curves of HS, IHS and GIHS parallel learning optimized BP neural network are compared under the same parameter settings as shown in Figure 1. Figure 1 shows that the three algorithms optimize the threshold and weight parameters of the neural network in parallel under the same parameter settings, and the GIHS algorithm achieves the optimal result when the number of iterations is less than 2000 The convergence speed and accuracy are better than the basic HS and IHS algorithms.

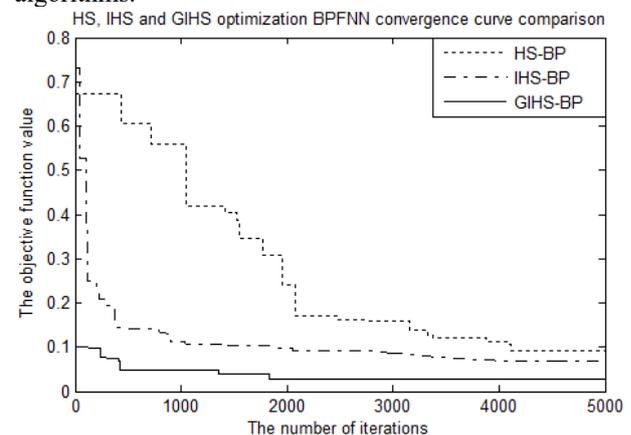


Fig.1 Comparison of HS and IHS and GIHS optimized BPFNN convergent curves

Prediction error comparison shown in Figure 2. The average classification error and average prediction error of 30 experiments were also calculated. As can be seen from Table 3 and Figure 2, the mean error of the GIHS-BP model in the weak classifier is the smallest, and the average error of the GIHS-BP_Adaboost model in the strong classifier is the least. The performance of the strong classifier is generally better than that of the predictive classifier Classifier performance is good.

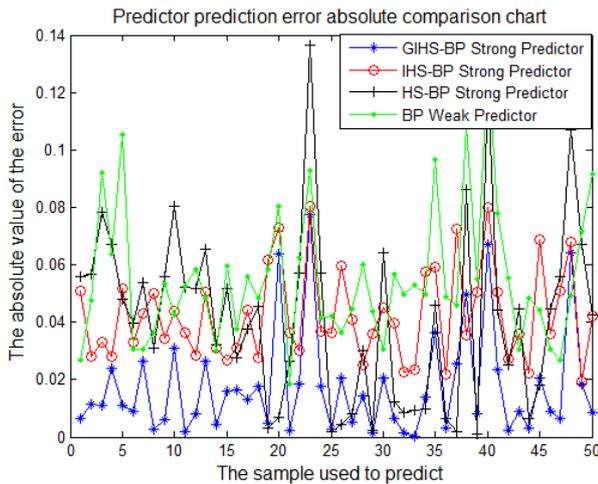


Fig.2 Comparison of prediction result error of models

Table 3 Prediction error Comparison of each model

	Classification Predictor Model	Average classification error	Average prediction error
Weak classifier and Predictor	BP	0.379 1	0.645 2
	HS-BP	0.197 3	0.318 6
	IHS-BP	0.174 5	0.237 8
	GIHS-BP	0.043 3	0.068 1
Strong classifier and Predictor	BP_Adaboost	0.232 0	0.352 2
	HS-BP_Adaboost	0.102 2	0.204 5
	IHS-BP_Adaboost	0.084 4	0.181 2
	GIHS-BP_Adaboost	0.036 7	0.026 3

5. Conclusion

First of all, based on the characteristics of harmonic search algorithm and its improvement status, this paper constructs a higher quality initial harmony library using good point set selection, and dynamically adjusts the three key parameters in the search process. At the same time, multiple iterations The new solution, make full use of harmony memory information to improve the global search ability and convergence speed of the algorithm.

Secondly, based on the improvement of the GIHS algorithm, aiming at the problems that the random initial weights and the thresholds of the BP neural network are apt to lead to the slow learning speed and easy to fall into the local solution, a BP neural network based on the GIHS algorithm Parallel integrated learning algorithm. This algorithm takes BP neural network as the basic framework, introduces GIHS algorithm in learning process, optimizes the network structure and connection weight of neural network by parallel interaction learning between GIHS algorithm and gradient descent method of BP neural network, in order to improve network convergence Speed and classification accuracy. Adaboost algorithm is used to integrate multiple GIHS-BP models

into a stronger classification prediction model.

Finally, the GIHS-BP_Adaboost model is applied to the company's financial alert problem. Analyze the factors that affect the financial status of the company, construct the company's financial early warning index system, and establish the financial early warning classification forecasting model based on this optimization algorithm. The simulation results based on the company's financial data show that the GIHS-BP_Adaboost model has higher prediction accuracy.

Acknowledgements

The National Natural Science Foundation of China under Grant No. 71601061; the key Natural Science Foundation of Education Department of Anhui Province under Grant No. KJ2017A624.

References

- Hao Qing. The Study on Financial Early-Warning for Listed Manufacturing Companies Based on BP Neural Network Optimized by Genetic Algorithm [D]. Hebei University,2014.6.
- Zhai Shen-yu. Research on Enterprise Financial Early Warning Of BP Neural Network Optimized By Modified HS algorithm.[D].JiLin University,2016.6.
- Zhan Jian-hua.The Research on Financial Early-Warning of Listed Companies Based on NRS-PSOBP Network[D].Shandong University of Financial and Economics,2012.4.
- Pan X T, Qu K Q. Wind Power Output Prediction with BP Neural Network Combining Genetic Algorithms[C]//Advanced Materials Research. Trans Tech Publications,2014, 860: 2526-252.
- Movahedi M, Kokabi A H, Reihani S M S, et al. Prediction and optimization of back-break and rock fragmentation using an artificial neural network and a bee colony algorithm[J].Bulletin of Engineering Geology & the Environment, 2016,75(1):27-36.
- Wang gai-ge, Guo Li-hong, Duan Hong. Target threat assessment using glowworm swarm optimization and BP neural network[J]. Journal of Jinlin University(Engineering and Technology Edition), 2013,43(4):1064-1069.
- Liu Qing-hua, Ding Wen-tao, Tu Juan-juan et.al. Improved BP_Adaboost Algorithm and its Application in Traffic Incident Detection[J]. JOURNAL OF TONGJI UNIVERSITY(NATURAL SCIENCE,2015,43(12):1829-1833.
- Geem Z W, Kim J H, Loganathan G V. A new heuristic optimization algorithm: Harmony search[J].Simulation,2001, 76(2): 60-68.
- Geem Z W. Optimal cost design of water distribution networks using harmony search[J]. Eng Optimiz, 2006,38(3): 259-280.
- Kang S L, Geem Z W. A new structural optimization

- method based on the harmony search algorithm[J].Comput Struct, 2004, 82(9/10): 781-798.
11. Kang S L, Geem Z W. A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice[J]. Computer Methods in Applied Mechanics and Engineering, 2005, 194: 3902-3933.
 12. Geem Z W, Tseng C, Park Y. Harmony search for generalized orienteering problem: best touring in China[J].Springer Lecture Notes in Computer Science, 2005, 3412:741-750.
 13. Zhou Ya-lan, Huang Tao. Modifications and Applications of Harmony Search Algorithm[J]. Computer Science,2014,41(6):52-57.
 14. Mahdavi M, Fesangharym,Damangir E. An improved harmony search algorithm for solving optimization problems[J].Applied Mathematics and Computation. 2007,188(2):1567-1579.
 15. Omran,Mahdavi M.Global-best harmony search[J].Applied Mathematics and Computation.2008,198(2):643-656.
 16. Shi L, Tang X, Lv J. PCA-based PSO-BP neural network optimization algorithm[C]//The 27th Chinese Control and Decision Conference (2015 CCDC). IEEE, 2015: 1720-1725.
 17. Wang Pei-chong, Li Li-rong, Gao Wen-chao,et.al. Hybrid Reverse Learning Artificial Fish School Algorithm Using Good Point Set [J]. Application Research of Computers, 2015, 32(7):1992-1995.
 18. Liu Xiang-pin, Xuan Shi-bin, Liu Feng. Artificial bee colony algorithm which introduces the good point set and the monkey group turn process [J]. Pattern Recognition and Artificial Intelligence, 2015, 28(1):80-89.