

Analysis and Research on Provincial Postgraduate Cultivation Performance Based on BP Neural Network

Shuai Wu^a, Peiyu Li, Xi Chen, and Hui Wang

School of Information Engineering, Henan University of Science and Technology, Henan, China.

^aheixue@live.com

Abstract. According to relevant data, the comprehensive competitiveness and quality of graduate education in Henan Province are in the middle of the 31 provinces in China, and its graduate education has a great room for improvement. Through the research on the current situation of educational performance evaluation at home and abroad, and the analysis of the limitations in the evaluation process of postgraduate training in Henan province, we proposed a performance evaluation method based on BP neural network. This method uses the theory and technology of information theory, analytic hierarchy process, and BP neural network to establish an educational performance evaluation model. Through this model, the shortcomings of the existing performance evaluation methods of graduate student training are overcome, and performance evaluation methods that are closer to the actual situation are excavated, which provides a scientific decision-making reference for the performance evaluation of graduate education in Henan Province. This experiment uses the university education data of Henan Province as the experimental data test set. The experimental results show that this model has good generalization ability and can provide reference for the reform and construction of postgraduate education in Henan province.

Keywords: Performance analysis, BP neural network, neighborhood mutual information, analytic hierarchy process, information entropy theory.

1. Introduction

At present, the main methods for the performance of graduate students at home and abroad are linear weight method, analytic hierarchy process, mathematical statistics method, fuzzy comprehensive evaluation method and other multi-index comprehensive evaluation methods. Most of these methods start from the perspective of overall scale, rather than evaluate performance from the perspective of input-output efficiency. This is not conducive to guiding the improvement of the utilization rate of educational resources, and it is also difficult to provide a decision-making basis for the optimal allocation of educational resources.

Data Envelopment Analysis (DEA) is a commonly used method to study the performance of multiple inputs and outputs [1-3]. This method attempts to comprehensively compare the input-output efficiency among universities, but only considers the overall efficiency, technical efficiency and scale efficiency of each decision-making unit. Without analyzing the specific reasons for the decision making unit falling into the invalid state from a single index, it cannot provide an effective basis for performance improvement.

In recent years, with the rapid development of education informatization, the number of postgraduate education data in our province has increased dramatically. These massive data have made it possible for us to adopt some new mining methods.

In view of the defects of traditional methods, as well as some current situations, and taking full account of various human and subjective factors, this paper introduces information theory into the mining process of performance analysis. Through the method of mutual information, BP neural network and hierarchical analysis, good mining results are obtained.

This research involves pedagogy, statistics, informatics, computer science and other multi-disciplinary research fields. It is a method of mining and predicting unknown performance evaluation results based on existing education data. It has a good reference role for graduate education evaluation in Henan Province.

2. Basic concepts and related principles

2.1 Analytic Hierarchy Process (AHP)

AHP was proposed by an American operations researcher and is currently relatively mature and effective in solving complex problems with multiple goals [4]. This is a typical systematic engineering method from qualitative analysis to quantitative analysis. In the process of graduate performance evaluation, this method can mathematicise people's complex thinking processes, quantify subjective judgments, and digitize various differences, which can effectively reduce the impact of human factors.

This article mainly uses the analytic hierarchy process to determine the weight of each output indicator used to determine performance. The main operation is divided into the following four steps:

Determine the index system

Through the comparison between the two indicators by experts, the comparison results are quantified with numbers from 1 to 9, and the average value is taken to form a judgment matrix

Table 1. Consistency index

Comparison results (compared between the two, the former is better than the latter)	Scaling
Equally important	1
Slightly stronger	3
Obviously important	5
Strongly important	7
Extremely important	9
2, 4, 6, 8 Between the front and back	

Check the consistency of the matrix. When the CR consistency ratio is less than 0.1, the degree of inconsistency is considered to be within the allowable range

Calculate the matrix to obtain the eigenvalues and eigenvectors satisfying the formula $AW = \lambda_{max}W$, and the components of W are the weights of the corresponding elements.

After the weight is obtained, the value used to measure performance can be obtained according to the weight and the index value.

2.2 Neighborhood Mutual Information (NMI)

Mutual information is a useful information measure in information theory, which can be used to measure the correlation between two sets of events. In information theory, the higher the entropy, the more information that can be transmitted, and the lower the entropy, the less information that can be transmitted[5].

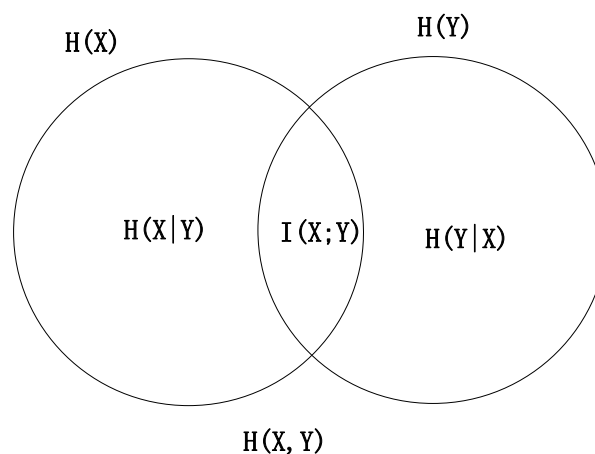


Fig. 1 Mutual information

At the same time, mutual information can be used to measure the degree of interdependence between two variables. In this project, the entropy principle in information theory is used to mine the correlation between input indicators to maximize the retention of input information

The mutual information calculation formula is as follows:

$$I(X; Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log\left(\frac{p(x, y)}{q(x)q(y)}\right)$$

Mutual information is a useful information measure in information theory, which can be used to measure the correlation between two sets of events. In information theory, the higher the entropy, the more information that can be transmitted, and the lower the entropy, the less information that can be transmitted.

At the same time, mutual information can be used to measure the degree of interdependence between two variables. In this project, the entropy principle in information theory is used to mine the correlation between input indicators to maximize the retention of input information

The mutual information calculation formula is as follows: When the variables X and Y are completely unrelated or independent of each other, the mutual information is the smallest, which means that the information entropy is the largest at this time, the overlapping information is the smallest, and the degree of dependence is the smallest.

Mutual information is generally used for discrete variables and cannot measure the nonlinear relationship between two variables. The education-related data used in the performance evaluation process is continuous data, so this paper chooses neighborhood mutual information to solve this problem.

In rough set theory, the neighborhood rough set model granulates the domain of the mixed information system by establishing neighborhood relations, so as to achieve an approximate approximation to the target knowledge [6-7].

The theorem of neighborhood mutual information is as follows:

Definition 1

Suppose that the set M is not empty. If $\forall x_a, x_b, x_k \in U$, there exists a uniquely certain real function P corresponding to it, and satisfies:

$$P(x_a, x_b) \geq 0 \text{ If and only if the condition } x_i = x_j \text{ is met, there is } P(x_a, x_b) = 0$$

$$P(x_a, x_b) = P(x_b, x_a)$$

$$P(x_a, x_k) < P(x_a, x_b) + P(x_b, x_k)$$

Then, P is the distance function on U. $\langle M, P \rangle$ is the metric space. In general, the commonly used distance functions are Manhattan distance and Euclidean distance.

Definition 2

Let $\langle M, P \rangle$ be non-empty, then $x \in M, \delta \geq 0$ is called a point set.

$$\delta(x) = \{y | P(x, y) \leq \delta, y \in U\}$$

is the neighborhood of x in the metric space $\langle M, P \rangle$

Definition 3

Given domain P, C refers to the conditional attribute describing P, and D is the decision-making attribute. According to C, a set of neighborhood relations R on the universe of discourse can be generated, and then $NDT \leq \langle P, R, D \rangle$ is called the neighborhood approximation space.

Definition 4

Given the universe of space $\{x_1, x_2, \dots, x_n\}$, conditional attributes C, $A \subseteq C, N_A$ are expressed as the neighborhood relations induced by feature subset A. Let x_i calculate the neighborhood based on A as $\delta_A(x_i)$, then the uncertainty of the neighborhood particle is:

$$NH_{\delta}^{x_i}(A) = -lb \frac{|\delta_A(x_i)|}{n}$$

The formula is based on 2. Correspondingly, the uncertainty of the neighborhood approximation space $\langle M, N_A \rangle$ is:

$$NH_{\delta}(A) = -\frac{1}{n} lb \frac{|\delta_A(x_i)|}{n}$$

Because of $\forall x_i, \delta_A \in U$ and $\frac{|\delta_A(x_i)|}{n} \leq 1, lb n \geq NH_{\delta}(A) \geq 0$ is obtained. If and only if $\forall x_i$ and $\delta_A(x_i) = 1$, there is $NH_{\delta} = lb n$. If and only if $\forall x_i$ and $\delta_A(x_i) = n$, there is $NH_{\delta} = 0$.

Definition 5

If and only if $A, B \in C$ are two sets of feature subsets describing the universe, and the neighborhood of x_i on $A \cup B$ is denoted as $\delta_{A \cup B}(x_i)$, then the joint entropy of A and B can be defined as:

$$NH_{\delta}(A, B) = -\frac{1}{n} \sum_{i=1}^n lb \frac{|\delta_{A \cup B}(x_i)|}{n}$$

When C is a decision attribute, $\delta_{A \cup B}(x_i) = \delta_A(x_i) \cap c_{x_i}$ can be obtained, at this time:

$$NH_{\delta}(A, C) = -\frac{1}{n} \sum_{i=1}^n lb \frac{|\delta_A(x_i) \cap c_{x_i}|}{n}$$

Definition 6

$A, B \subseteq C$ is the two sets of feature subsets describing the universe. Given feature subset A, the neighborhood conditional entropy of B relative to A is:

$$NH_{\delta}(B|A) = -\frac{1}{n} \sum_{i=1}^n lb \frac{|\delta_{A \cup B}(x_i)|}{|\delta_A(x_i)|}$$

Definition 7

$A, B \subseteq C$ is the two sets of feature subsets describing the universe. The neighborhood mutual information of A and B can be defined as:

$$NMI_{\delta}(A; B) = -\frac{1}{n} \sum_{i=1}^n lb \frac{|\delta_A(x_i)| \cdot |\delta_B(x_i)|}{n |\delta_{A \cup B}(x_i)|}$$

2.3 BP neural network

Neural network is a technology that abstracts and simulates the general characteristics of the human brain or nature. In recent years, the application of neural network in the field of prediction and empirical research has achieved remarkable results. BP neural network has a fairly strong ability to process information, and has self-learning and self-adaptive capabilities in the process of processing data, which can better explain the non-linear characteristics contained in the data.[8]

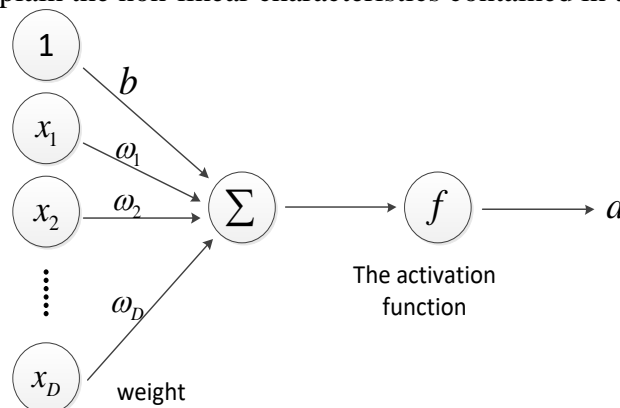


Fig. 2 Neurons

Neurons are the basic units that form neural networks. In a biological neural network, each neuron is connected to other neurons. When it is "excited", it will send chemicals to the connected neuron to transmit information to the next neuron. In an artificial neural network, a neuron receives input signals from n other neurons. These input signals are transmitted through a weighted connection. The total input value received by the neuron will be compared with the neuron threshold, and then Through the "activation function" processing to generate neuron output.

Backpropagation neural network is a three-layer feedforward network, including input layer, hidden layer, and output layer. For the input signal, it must first propagate forward to the hidden node, and then pass the output information of the hidden node to the output node after the function action, and finally get the output variable result.[9]

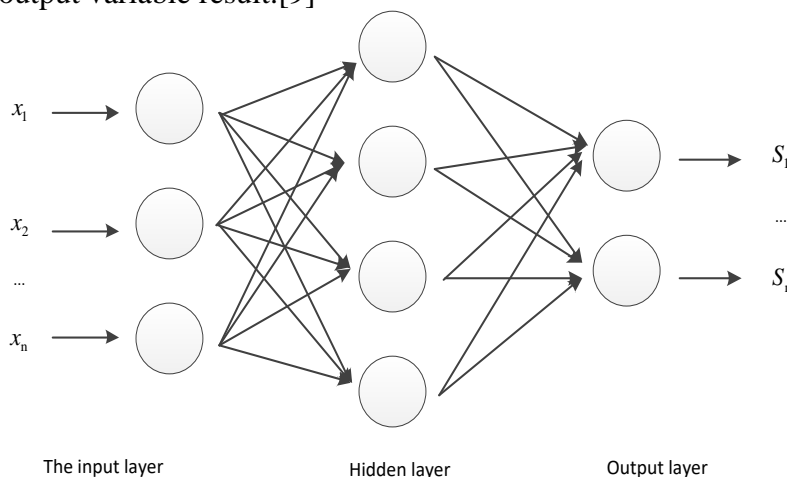


Fig. 3 Neural Network

The learning process of the algorithm consists of forward propagation and back propagation. The input information is processed by the hidden layer unit and passed to the output layer. If the desired output cannot be obtained in the output layer, it enters back propagation and modifies the weight of the neuron. Make the error signal reduce. The system continuously loops these two processes, repeating the learning, until the error between the output value and the expected value is reduced to the specified range, the system stops learning.

In this experiment, pytouch is used to realize the related calculation of BP neural network. pytouch is Facebook's open source machine learning library. It is a python-based sustainable calculation package with powerful GPU accelerated tensor calculation, which can effectively improve the calculation speed.

3. Selection of evaluation indicators

In the performance analysis process, the input and output indicators are mainly based on the data available in the current process of college graduate training. In the initial selection, it is necessary to include as much as possible the statistical index data.

3.1 Confirmation of input indicators

Combining the data characteristics of colleges and universities in our province and the status quo of postgraduate training, the selected investment is mainly divided into three aspects, capital investment, manpower investment and equipment investment. They are: number of teachers, teaching volume, doctoral degree teachers, full-time teachers, project funding, unit talent training personnel funding, unit scientific research personnel funding, teaching staff funding, unit talent training room area, unit room area, unit science Research room area, teacher team room area, unit personnel training equipment investment, unit scientific research equipment investment, faculty equipment investment, unit scientific research total, unit equipment funding, key laboratory, natural science fund, etc. 19 inputs Indicators are used as raw input data.

After the dimensionless normalization of the indicator data, in order to avoid excessive substitution between the selected indicators, the information theory is used to analyze the relevance of the input indicators.

According to the principle of information theory, mutual information can measure the mutual relationship between two variables, and can measure the amount of information between message sets. This method can reduce the dimensionality while preserving the overall characteristics as much as possible.

Calculate the information entropy of 19 input indicators, then delete individual indicators in turn, calculate and compare the information entropy of the remaining indicators. Compare and observe the calculation results, delete the index items that have the least impact on the entropy value after deletion, and then cycle.

The information entropy value change after each deletion is shown in Figure 4.

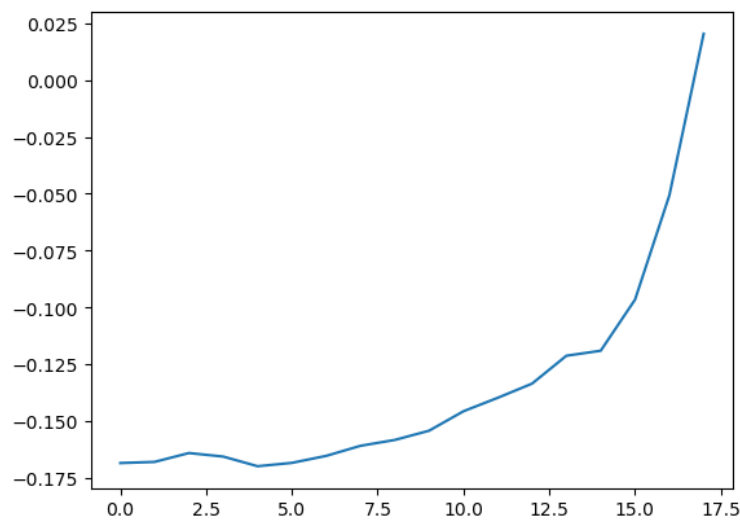


Fig. 4 Entropy Value

It can be seen from the figure that when the tenth indicator is deleted, the entropy value has changed significantly, and when the first 10 indicators are deleted, the amount of information has little effect.

Therefore, the input indicators selected in this article are: Natural Science Foundation, Key Laboratory, Unit Scientific Research, Unit Personnel Training Area, Unit Personnel Training Personnel Expenses, Teaching Staff Housing Area, Teaching Staff Expenses, Unit Subject Research Staff funding, number of PhD teachers.

3.2 Determination of output indicators

The output data indicators mainly include the following 7: intellectual property rights and works, employment rate, graduation rate, number of students, journal papers, competition awards, and outstanding alumni.

Here, we use the analytic hierarchy process. The analytic hierarchy process quantifies people's subjective evaluation of performance, this complex system of thinking process, and helps people maintain the consistency of their thinking models, thereby reducing the influence of factors that are considered.

Find 10 experienced university professors to set the output index weights, and assign values according to the comparison of the two indexes according to the 1-9 scale method, and then discuss the results. The final result is shown in the Figure6.

Check the consistency of the matrix. When the CR consistency ratio is less than 0.1, the degree of inconsistency is considered to be within the allowable range. After testing, the judgment matrix can pass the consistency test.

Table 2. Index Value

	Books	Job_rate	Graduate_rage	Num	Article	Award	Outstanding_graduate
Books	1	1/2	8	3	6	3	1/2
Job_rate	2	1	7	8	5	4	1
Graduate_rage	1/8	1/7	1	3	1/3	1/5	1/8
Num	1/3	1/8	1/3	1	1/5	1/7	1/8
Article	1/6	1/5	3	5	1	1/2	1/6
Award	1/3	1/4	5	7	2	1	1/5
Outstanding_graduate	2	1	8	8	6	5	1

After each group of data is normalized, multiplied by the weight, the final measurement value for performance evaluation can be obtained.

According to the metric value obtained in the final calculation, the result value is evaluated according to the following grades, among which the good and the difference are represented by the numbers 3, 2, 1, and 0, and are used as the output result of the system.

Table 3. Performance evaluation grade

Performance evaluation grade	Scaling
excellent	0.75-1
good	0.5-0.75
medium	0.25-0.5
difference	0-0.25

After each group of data is normalized, multiplied by the weight, the final measurement value for performance evaluation can be obtained.

4. Construction of bp neural network performance index evaluation model

4.1 Abbreviations and Acronyms

The experimental data were obtained from universities in Henan Province and the network , and the experimental data were divided into training set and test set.

The experimental environment is Windows10 + Anaconda + Mysql + Pytoch + VisualStudio. Anaconda provides a Python environment for experiments , while PyTouch provides library functions for training BP neural networks.

4.2 Input layer

Using the neighborhood mutual information interest rate , the selected input indexes were reduced from 19 to 9 after correlation analysis .

Table 4. Index Input

number	Input indicators
1	Natural science fund
2	Key laboratories
3	Total amount of scientific research
4	Room area for talent training
5	Funds for personnel training
6	Room area for teachers
7	Faculty staff funds
8	Funding for discipline researchers
9	Number of teachers with doctoral degrees

4.3 Hidden layer

In the training process of BP neural network, the number of hidden nodes can affect the performance of the training model. At present, there is no definite calculation formula for the number of hidden layer nodes. According to the traditional empirical formula

$$h = \sqrt{m + n} + a (0 < a < 10)$$

The approximate reasonable range can be calculated.

Where, H is the number of nodes in the hidden layer, M is the number of nodes in the input layer, and N is the number of nodes in the output layer. A is the regulation constant between 1 and 10.

In this paper, $m=9$, $n=4$, the number range of nodes in the hidden layer can be determined as [3, 13].

Through experimental comparison and comprehensive consideration of factors such as time cost, iteration number and error size of machine learning, the number of hidden layer nodes is set as 8.

4.4 Output layer

According to the performance evaluation system of domestic graduate students, it is divided into four levels: excellent, good, medium and poor, so the number of output neurons is set as 4

The activation function of the hidden layer adopts the ReLU function, and the step size of the training process is 0.02

4.5 Training model

The selected input index data and output index data are de-dimensionalized as the training data of BP neural network model.

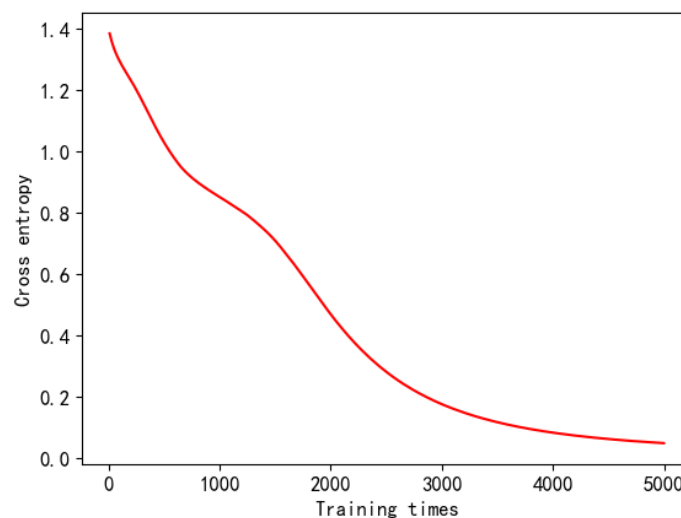


Fig. 5 Cross entropy

Cross entropy can be used to determine the degree of proximity between the actual output and the expected output. The smaller the value of cross entropy is, the closer the two probability distributions are.

As shown in the figure above, after 3000 times of training, the cross entropy change is no longer obvious.

The trained model is saved for further verification.

4.6 Validation

The trained model is verified by the validation set, and the results show that the prediction accuracy is over 90%, which is in line with the expectation.

5. Conclusion

This paper conducts a comprehensive study on the performance evaluation of colleges and universities in Henan Province, aiming to analyze the current situation and influencing factors of the performance of postgraduate training in colleges and universities in Henan Province, and carries out simulation on relevant data through machine learning model BP neural network to verify the generalization ability.

The weight of performance related indicators was determined by the chromatography analysis method, and the input indicators were screened by the information entropy theory, so as to effectively avoid the influence of human subjective factors and reasonably simplify the input indicators in a more scientific way.

BP neural network has the ability of self-learning and self-adaptation, which reduces the influence of human factors. Increases the credibility of the results.

Due to limited data sources and incomplete data of some colleges and universities, the selection of qualitative indicators has certain limitations. In the follow-up research, with the improvement of databases of colleges and universities, this evaluation model can also achieve better results.

References

- [1] Boon L. Lee, Andrew C. Worthington. A network DEA quantity and quality-orientated production model: An application to Australian university research services. 2016, 60:26-33.
- [2] SUEYOSHI T. Neural computation for DEA performance analysis [J]. *Asia Pacific Management Review*, 2005, 10(6):381–389.
- [3] EMROUZNEJAD A, SHALE E. A combined neural network and DEA for measuring efficiency of large-scale datasets[J]. *Computers & Industrial Engineering*, 2008, 56(1): 249-254
- [4] T L SAATY. The Analytic Hierarchy Process [M]. New York: McGraw – Hill, Inc, 1980: 51-62
- [5] Battiti R. Using mutual information for selecting features in supervised neural net learning. [J]. *IEEE transactions on neural networks*, 1994, 5(4).
- [6] Qinghua Hu, Lei Zhang, David Zhang, Wei Pan, Shuang An, Witold Pedrycz. Measuring relevance between discrete and continuous features based on neighborhood mutual information[J]. *Expert Systems with Applications*, 2011, 38(9).
- [7] Qinghua Hu, Wei Pan, Shuang An, Peijun Ma, Jinmao Wei. An efficient gene selection technique for cancer recognition based on neighborhood mutual information[J]. *International Journal of Machine Learning and Cybernetics*, 2010, 1(1).
- [8] Bao Long Hu, Ji Ren Xu, Huai Hui Gao, Ji Hai Liu, Ke Ren Wang. Modified BP Neural Network Model is Used for Odd-Even Discrimination of Integer Number[J]. *Applied Mechanics and Materials*, 2013, 2746.
- [9] Ying Shu, Guang-hua Xu. Multi-level Dynamic Fuzzy Evaluation and BP Neural Network Method for Performance Evaluation of Chinese Private Enterprises[J]. *Wireless Personal Communications*, 2018, 102 (4).