Scholars Journal of Physics, Mathematics and Statistics

OPEN ACCESS

Abbreviated Key Title: Sch J Phys Math Stat
ISSN 2393-8056 (Print) | ISSN 2393-8064 (Online)
Journal homepage: https://saspublishers.com/sjpms/

Comprehensive Evaluation Model of Infectious Disease Epidemic Degree Based on PCA and BP Neural Network

Wang Zhili[1*], Ding Xuanyi[2]

[1]College of Applied Mathematics, Chengdu University of Information Technology, Chengdu 610225, Sichuan, P.R. China
[2]College of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, Sichuan, P.R. China

DOI: 10.36347/sjpms.2020.v07i05.002 | **Received:** 11.05.2020 | **Accepted:** 18.05.2020 | **Published:** 19.05.2020

***Corresponding author:** Wang Zhili

Abstract **Original Research Article**

Aiming at the complexity and correlation characteristics of infectious disease prevalence evaluation factors, a comprehensive evaluation method of infectious disease prevalence combined with principal component analysis and BP neural network is proposed. After comprehensive analysis, the evaluation index system based on the population, the number of infections, the number of deaths and deaths, the duration of the epidemic, the economic status, medical conditions, population density, epidemic prevention policies, and the number of infected countries is selected. The relevant data of "epidemic" and "pandemic" constitutes a BP neural network evaluation model for evaluating the prevalence of infectious diseases. The research results show that the combined method of PCA and BP neural network reduces the input variables from 9 to 2, avoiding the influence caused by the correlation of variables, simplifying the evaluation process, and the results are more reasonable. The actual results and the calculation results of the comprehensive evaluation model of the prevalence of infectious diseases based on PCA and BP neural network support each other.

Keywords: Infectious diseases; Prevalence; Principal component analysis; BP Neural Network.

INTRODUCTION

On March 12, 2020, the World Health Organization (WHO) announced that the viral pneumonia (COVID-19) caused by the coronavirus sweeping the world is a pandemic. Although SARS affects 26 countries, it is still not considered a pandemic, and MERS is not considered a pandemic [1]. The WHO says the pandemic is "a global spread of new diseases." Factors that affect the prevalence of infectious diseases include population, number of infected countries, number of infections, death toll, duration of epidemic situation, economic status, medical conditions, population density, epidemic prevention policy, number of infected countries, etc. There is no strict standard for quantifying the pandemic or not, and there is no threshold for the number of cases or deaths that trigger this definition. Therefore, the establishment of an evaluation model for the prevalence of infectious diseases is very important. There are many traditional evaluation models. Such as fuzzy comprehensive evaluation method [2], principal component analysis method [3] analytic hierarchy process [4], safety checklist [5], etc. Most of these evaluation models are linear models. In the actual evaluation process, when evaluating a complex epidemic system, it is difficult to fully grasp the evaluation index variables, and there are uncertain factors such as variables affecting and intersecting with each other, which leads to the nonlinearity of the evaluation process (function mapping). The BP neural network is a nonlinear system composed of a large number of simple processing units. It has the characteristics of high nonlinearity, self-learning, self-organization and effective function approximation. It is suitable for the safety evaluation of nonlinear complex systems [6]. However, too many evaluation indexes will cause the traditional BP neural network to have problems such as slower convergence speed, reduced network performance and calculation accuracy, and rapid increase in calculation time.

Therefore, we can construct a non-linear comprehensive evaluation model of the prevalence of infectious diseases based on PCA and BP neural network. First, select the population, the number of infected countries, the number of infections, the number of deaths and deaths, the duration of the epidemic, economic status, medical conditions, population

density, epidemic prevention policies, and the number of infected countries as the evaluation indicators of the pandemic, and quantify the qualitative indicators. Then the principal component analysis method is used to process the evaluation indicators, forming a new indicator system, which effectively eliminates the correlation between the original indicators. Reduced the input dimension of BP neural network. Then select the appropriate principal component as the input layer node index of the BP neural network, to reduce the dimension of the newly learned sample space, improve the operation efficiency and accuracy, and construct the network topology. Finally, the output value of BP neural network can define epidemic and pandemic. Finally, a set of data was selected to compare the results of the model and test the rationality of the model.

Comprehensive Evaluation Model of Infectious Disease Epidemic Degree Based on PCA and BP Neural Network

PCA

Principal component analysis as a basic mathematical analysis method, its practical applications are very wide, such as demography, quantitative geography, molecular dynamics simulation, mathematical modeling, mathematical analysis and other disciplines have been applied, is a commonly used multivariate Analysis method [7]. When using statistical analysis methods to study multi-variable topics, too many variables will increase the complexity of the topic. Principal component analysis is to delete the redundant variables (closely related variables) for all the variables originally proposed, and establish as few new variables as possible, so that these new variables are irrelevant, and these new variables are reflecting The information aspect of the subject should keep the original information as much as possible. Try to recombine the original variables into a new set of independent comprehensive variables. At the same time, according to the actual needs, you can take out a few less comprehensive variables to reflect the information of the original variables as much as possible. Principal component analysis is also a method used for dimensionality reduction in mathematics [8]. The main steps are as follows:

There are p evaluation indicators $X_1, X_2, \cdots X_p$ and p evaluation indicators of n evaluation objects forming the original data matrix X^*.

1) Normalize the original data matrix X^* to eliminate the influence of dimension and the difference in order of magnitude to obtain a standardized data matrix X.

2) Establish the correlation coefficient matrix $R = (r_{ij})_{np}$ of the standardized data matrix X. Where r_{ij} is the correlation coefficient between index X_i and index X_j.

3) Calculate the eigenvalue $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p > 0$ of the correlation matrix R, and its corresponding eigenvector $u_1, u_2, \cdots u_p$,

Where $u_i = (u_{i1}, u_{i2}, \cdots u_{ip})(i = 1,2,\cdots, p)$.

3) Thereby obtaining p principal component $Y_1, Y_2, \cdots, Y_p$ and Y_i is a linear combination of the variable $X_1, X_2, \cdots X_p$, which is

$$Y_i = u_{i1}X_1 + u_{i2}X_2 + \cdots u_{ip}X_p \quad i = 1,2,\cdots, p$$

The eigenvalue λ_i is the variance, which reflects the role of the first i principal component in describing the evaluated object. The larger λ_i, the greater the contribution to the total variation, and the contribution rate is $a_i = \lambda_i / \sum_k \lambda_k$.

4) Select the principal components according to the research accuracy, usually select the smallest integer m of 85% that makes the cumulative contribution rate $\sum_i a_i$ of the principal components exceed, and finally determine the first m principal components $Y_1, Y_2, \cdots, Y_p$, and use the m principal components as comprehensive indicators to replace the original p evaluation indicators $X_1, X_2, \cdots X_p$, get a new index system.

BP BP neural network

Basic principles of BP model

The learning process consists of two processes: the forward propagation of the signal and the backward propagation of the error. During forward propagation, the mode acts on the input layer. After the hidden layer is processed, the incoming error is propagated in the backward propagation stage, and the output error is sorted according to a certain factor. In the form, the hidden layer returns to the input layer layer by layer, and "distributes" to all units of each layer, so as to obtain the reference error or error signal of each layer unit as a basis for modifying the weight of each unit. The weight is continuously modified Is the process of network learning. This process continues until the error of the network output gradually decreases to an acceptable level or reaches the set number of learning times [9].

Training of BP neural network

The BP algorithm obtains this kind of input and output by the event of "training", and the appropriate linear or nonlinear relationship between the outputs. The process of "training" can be divided into two stages: forward transmission and backward transmission:

[1] Forward transmission stage:

(1)Take a sample P_i, Q_j from the sample set and enter P_i into the network;

(2)Calculate the error measure E_1 and the actual output $O_i = F_L(...(F_2(F_1(P_i W^{(1)})W^{(2)})...)W^{(L)})$;

(3)Make an adjustment to the weight value $W^{(1)}, W^{(2)}, ...W^L$, and repeat this cycleuntil $\sum E_i < \varepsilon$.

[2] Backward propagation stage-error propagation stage：

(1) Calculate the difference between actual output O_p and ideal output Q_i;

(2)Adjust the weight matrix of the output layer with the error of the output layer;

(3) $E_i = \frac{1}{2}\sum_{j=1}^{m}(Q_{ij} - O_{ij})^2$;

(4)Use this error to estimate the error of the direct leading layer of the output layer, and then use the error of the output layer leading layer to estimate the error of the previous layer. In this way, the error estimates of all other layers are obtained;

(5)And use these estimates to modify the weight matrix. Form a process that passes the error shown at the output to the output in stages in the opposite direction to the output signal.

The error measure of the network about the entire sample set: $E = \sum_i E_i$.

Comprehensive Evaluation Model of Infectious Disease Epidemic Degree Based on PCA and BP Neural Network

When evaluating the wide-ranging capacity of infectious diseases, it is necessary to find out the factors that affect the evaluation object. There are many variables that affect the safety of complex systems. Each variable reflects system information from different aspects and has certain relevance, so that the information it reflects overlaps to a certain extent, so it is not appropriate to directly use BP neural network for evaluation.

The nonlinear comprehensive evaluation model based on principal component analysis (PCA) and BP neural network proposed in this paper uses the data dimensionality reduction function of principal component analysis to extract the feature of the input layer node index of the neural network, and according to the principal component judgment standard Select the appropriate principal component as the new network input layer node index, in order to reduce the dimension of the new learning sample space, improve the efficiency and accuracy of the operation. The evaluation model of comprehensive BP neural network based on principal component analysis mainly includes the following steps.

(1) Establish an evaluation index system. According to the actual situation of the prevalence of infectious diseases, the index system of the evaluated system is determined according to the evaluation objectives.

(2) Collect sample data and process sample data. Make use of the preparation of the qualitative index evaluation table, and talk about the quantification of qualitative indexes. Then normalize it to eliminate dimensional effects.

(3) The principal component analysis method in SPSS is used to extract the feature of the original sample data, eliminate the coupling relationship between the indicators, and determine the number of principal components and the correlation coefficient matrix between the original variables and the principal components according to the principal component judgment standard.

(4) Using the correlation coefficient matrix, the original sample data is converted into standardized evaluation data expressed by principal components, and used as the sample data of the new input layer of the BP neural network model, and the output is the evaluation target value (expected value). The number of selected principal components is the number of nodes in the input layer of the integrated BP neural network model; the number of hidden layers and the number of neurons in each layer are determined by experiments or stepwise growth methods, and then the topology of the neural network model is determined .

(5) Use the converted standardized evaluation data to train and learn the integrated BP neural network model, so that the mold plow can obtain sufficient knowledge and experience.

(6) Input the corresponding index value of the evaluation object into the trained BP neural network model to obtain the evaluation result.

Data processing

1. Quantify qualitative data
2. The indicators in the collected data are missing and will not be considered
3. Randomly extract the collected data
4. Normalize the data to eliminate the dimension

Quantification of qualitative indicators:

Table-1: Classification of economic conditions

Economic status grade	Poor	Ordinary	Good	Splendidly
Mathematical expression	0.25	0.5	0.75	1

Table-2: Classification of medical conditions

Medical condition level	Poor	Ordinary	Good	Splendidly
Mathematical expression	0.25	0.5	0.75	1

Table-3: Classification of population density

Population density grade	Small	Medium	Big	Huge
Mathematical expression	0.25	0.5	0.75	1

Table-4: Classification of defense policies

Defensive policy level	Poor	Ordinary	Good	Splendidly
Mathematical expression	0.25	0.5	0.75	1

Table -5: Infectious disease prevalence capacity classification

level	Mathematical expression	Degree
S_1	$(1,0)$	Pandemic
S_2	$(0,1)$	Epidemic

Model solution

The 9 indicators we selected are based on the data after standardization, using the correlation coefficient matrix $R=(r_{ij})_{np}$ and corresponding feature vector $u_1,u_2,\cdots u_p$ obtained by the spss statistical software, and the contribution rate $a_i=\lambda_i/\sum_k \lambda_k$, and according to the principle that the cumulative contribution rate exceeds 85% of the smallest positive number from 9 Two main components are extracted from each component. As shown in Table 6 and Table 7 below.

Table-6: Eigenvalues of principal components and variance contribution rate

main ingredient	Eigenvalues	Contribution rate	Cumulative contribution rate
Y_1	5.609	62.323	62.323
Y_2	2.142	23.795	86.120

Table-7: Factor load matrix of principal components

Index	Main layer	
	Y_1	Y_2
X_1	0.953	-0.227
X_2	-0.028	0.921
X_3	-0.422	0.821
X_4	-0.91	-0.128
X_5	-0.638	-0.558
X_6	0.982	0.17
X_7	0.977	-0.204
X_8	0.933	-0.212
X_9	0.827	-0.393

The main index is used to extract the characteristics of the original index, the original index system ($X_1,X_2,X_3,\cdots,X_8,X_9$) is simplified, and two main comprehensive indexes (Y_1,Y_2) are obtained. According to the

data in Table 7, we can use linear transformation to get the relationship between the two principal components Y_1, Y_2 and the index variable $X_1, X_2, X_3, \cdots, X_8, X_9$ as:

$$Y_1 = 0.953X_1 - 0.028X_2 - 0.422X_3 + \cdots + 0.933X_8 + 0.827X_9$$

$$Y_2 = -0.227X_1 + 0.921X_2 + 0.824X_3 + \cdots - 0.212X_8 - 0.393X_9$$

We bring the survey data $X_1, X_2, X_3, \cdots, X_8, X_9$ into the relationship to get a new principal component Y_1, Y_2, and then we bring the epidemic degree of infectious diseases to the corresponding epidemic, see Table 8.

Table-8: data

Category	Y_1	Y_2	level	Degree
Black Death	-1.63399	-0.72733	S_1	Pandemic
Spanish influenza	-0.67428	1.61787	S_1	Pandemic
SARS	0.47941	-0.27306	S_2	Epidemic
MERS	0.03244	-0.87944	S_2	Epidemic
H1N1	0.77378	0.82783	S_1	Pandemic
Ebola	1.02264	-0.56586	To be tested	To be tested

These data constitute the learning sample of the BP neural network. We will bring each index value of Ebola into the relationship to obtain a new principal component as an input into the learned BP neural network to predict its value. It can be determined whether Ebola is an epidemic or a pandemic, and then compared with the actual results. We compare the calculated value with the actual value, as shown in Table 9 below

Table-9: Comparison Table

Name	Mathematical expression	level	Degree
Training output	$(0,1)$	S_2	Epidemic
Actual value	(0.083,0.9532)	S_2	Epidemic

Considering that the neural network model can also be used for prediction, the neural network model is compared with the non-linear comprehensive evaluation model based on principal component analysis and BP neural network to obtain Table 11.

Table-10: Error comparison table

Name	Mathematical expression	level	Degree	Error
BP neural network model	(0.1862,0.8544)	S_2	Epidemic	0.3318
Comprehensive evaluation model	(0.083,0.9532)	S_2	Epidemic	0.1298

It can be seen from Table 10 that both the pure neural network model and the non-linear comprehensive evaluation model based on principal component analysis and BP neural network can make predictions. Principal component analysis and the nonlinear comprehensive evaluation model of BP neural network predict the error of the predicted value is much larger, which shows that the nonlinear comprehensive evaluation model based on principal component analysis and BP neural network is more feasible and better Define epidemics and pandemics.

CONCLUSIONS

The non-linear properties of the comprehensive evaluation model based on principal component analysis and BP neural network proposed by us are closer to the actual evaluation process, and make up for the shortcomings caused by the overlapping of information between the original variables of the traditional linear evaluation method. By using the principal component analysis method to extract the feature data of the index data, the main comprehensive index is obtained, which reduces the dimension of the new learning sample and reduces the correlation

between the components, and improves the convergence speed of the network and the efficiency of learning and training. Finally, for the complex evaluation of the influence of infectious diseases, more ideal results are given, and the feasibility and effectiveness of the model are illustrated by examples. Compared with other general evaluation models, the error is smaller.

REFERENCES

1. Lin Siyu, Fang Pengqian, Yao Yao, Li Lu. Research on the cooperative mechanism of county-level multi-sector response to influenza pandemic [J]. Medicine and Society. 2009; 22 (10): 14-15.
2. Xie Meili, Zhang Xinhuan, Wu Jinhong, Wang Juan. Passenger satisfaction evaluation of Hangzhou Metro based on fuzzy comprehensive evaluation method [J]. Transportation and Transportation. 2020; 36(02): 78-81.
3. Feng Qiang. Evaluation of informatization teaching in higher vocational mathematics courses based on principal component analysis model [J]. Modern Information Technology. 2019; 3 (23): 101-103.
4. Chen Wei, Liu Xuejiao, Xia Yingjie. Multi-factor reputation evaluation model of IoV based on AHP [J]. Journal of Zhejiang University (Engineering Science Edition). 2020;54(04): 722-731.
5. Zhang Qiguo, Zhang Zhongjian, Ding Jixin. Quantitative evaluation of the current safety status of tailings pond based on improved safety checklist method [J]. Safety and Environmental Engineering. 2018, 25 (05): 121-126.
6. Bai Baoguang, Fan Qingxiu, Zhu Honglei. Research on Public Service Quality Evaluation Model of High-tech Zone Based on BP Neural Network [J]. Mathematics Practice and Cognition. 2020, 50(03): 154-163.
7. Lin Haiming, Du Zifang. Problems that should be paid attention to in the comprehensive evaluation of principal component analysis [J]. Statistical Research. 2013, 30(08): 25-31
8. Han Xiaodong, Zhang Yaohui, Sun Fujun, Wang Shaohua. Index weight determination method based on principal component analysis [J]. Journal of Sichuan Military Engineering. 2012, 33 (10): 124-126
9. Phase One Technology Product R & D Center, MATLAB6.5 assisted neural network analysis and design [M]., Beijing: Electronic Industry Press; 2003.