MEDS Clinical Medicine (2021) 2: 77-81
Clausius Scientific Press, Canada

DOI: 10.23977/medsc.2021.020215
ISSN 2616-1907

Strategy for the Protection of Global EDP Refugees

Tong Liu, Yijie Huang, Bingxue Li

College of Marine Electrical Engineering, Dalian Maritime University, Dalian, Liaoning, 116026

Keywords: Principal component Analysis, BP Neural Network, Risk Assessment Model

Abstract: In recent years, the situation of EDP refugees has become increasingly serious. In order to determine the scope of EDP refugees, we mainly forecast their current and future numbers and assess the risk of cultural loss of EDP refugees themselves. First of all, we select 10 factors that affect the number of EDP, and use **Principal component Analysis** to screen out 6 main factors. Next, a **BP Neural network** model is established, which takes six main factors as network inputs to predict the number of EDP refugees in 2020 and 2021. Then we set up a **Risk Assessment Model** to assess the risk of cultural loss of EDP refugees, looking for six main factors that affect the risk of cultural loss, and giving different scores to different factors and different situations, and finally divided into four levels of cultural loss risk. This model can be applied to all EDP refugees to assess the risk level of their own cultural loss, and then take corresponding measures to protect their own culture.

1. Introduction

The scope of the problem of EDP refugees is of great concern to the United Nations and the ICM-F, so we need to establish models to predict the current and future number of EDP refugees and to assess the risk of loss of EDP refugee culture.

2. Establish the prediction model and solve

2.1 Principal component Analysis [1]

We need to build a model to predict the current and future EDP population, and we have identified 10 factors affecting EDP population, namely, earthquake, flood, drought, tropical cyclone, desertification, avalanche, hill fire, volcanic eruption, tsunami, sea level rise. Through the website [2] to find out the number of EDP affected by these ten factors, and then through the formula to determine several principal components.

There are 10 index variables for principal component analysis. They are $x_1, x_2, \ldots, x_{10}$,There are 9 subjects to be evaluated, The value of the j-th index of the i-th evaluation object is a_{ij}, Convert each index value a_{ij} into a standardized index value $\overline{a_{ij}}$

$$\overline{a_{ij}}=\frac{a_{ij}-u_j}{s_j},\ i=1,2,\ldots,9;\ j=1,2,\ldots,10 \tag{1}$$

In the formula: $u_j=\frac{1}{9}\sum_{i=1}^{9} a_{ij}$; $s_j=\sqrt{\frac{1}{8}\sum_{i=1}^{9}(a_{ij}-u_j)^2}$, j=1, 2, …10

✧ Calculate the correlation coefficient matrix R.

Correlation coefficient matrix $R=(r_{ij})_{m\times m}$,

$$r_{ij}=\frac{\sum_{k=1}^{n}\overline{a_{k1}\cdot a_{kj}}}{8},\ i,\ j=1,\ 2,\ \ldots 10 \tag{2}$$

In the formula: $r_{ii}=1$, $r_{ij}=r_{ji}$, r_{ij} is the correlation coefficient between index i and index j.

✧ Calculate eigenvalues and eigenvectors.

Calculating the eigenvalue of the correlation coefficient matrix R, $\lambda_1 \geq \lambda_2 \geq \cdots \lambda_{10} \geq 0$, And corresponding feature vectors $u_1, u_2, \ldots, u_{10}$, among them $u_j = [u_{1j}, u_{2j}, \ldots, u_{10j}]^T$, Ten new index variables are composed of feature vectors:

$$y_1 = u_{11}\bar{x}_1 + u_{21}\bar{x}_2 + \cdots + u_{10\,1}\bar{x}_{10},$$
$$\vdots$$
$$y_{10} = u_{1\,10}\bar{x}_1 + u_{2\,10}\bar{x}_2 + \cdots + u_{10\,10}\bar{x}_{10},$$

✧ Select p(p≤10) principal components and calculate the comprehensive evaluation value.

① Calculate the eigenvalue λ_j (j = 1, 2,...,10) information contribution rate and cumulative contribution rate.

$$b_j = \frac{\lambda_j}{\sum_{k=1}^{10}\lambda_k}, j=1,2,\ldots,\ 10 \tag{3}$$

Is main component y_j information contribution rate, at the same time,

$$a_p = \frac{\sum_{k=1}^{p}\lambda_k}{\sum_{k=1}^{10}\lambda_k} \tag{4}$$

Is the cumulative contribution rate of principal component y_1, y_2,…, y_p. When a_p close to 1(Usually take $a_p = 0.85,\ 0.90,\ 0.95$), Then select the first p index variables y_1, y_2,…, y_p as p principal components, Instead of the original 10 index variables, p principal components can be comprehensively analyzed.

② Calculate the comprehensive score

$$Z=\sum_{j=1}^{p} b_j y_j, \tag{5}$$

In the formula: b_j is the information contribution rate of the j-th principal component, which can be evaluated based on the comprehensive score.

✧ It is calculated through matlab that the characteristic roots of earthquake, flood, sea level rise, tsunami, avalanche, volcanic eruption, mountain fire, desertification, drought and tropical cyclone are: 9.165, 0.7279, 0.0760, 0.0174, 0.0089, 0.0032, 0.0014, 0.0001, 0.0001, 0.0001; Contributions are: 91.6500%, 7.2790%, 0.7596%, 0.1741%, 0.0895%, 0.0320%, 0.0120%, 0.0010%, 0.0010%, 0.0010%; The cumulative contribution of the first six characteristic roots has reached 99.9842%. Therefore, the first six main factors can be used instead of the ten influencing factors.

2.2 BP neural network algorithm [3]

➢ Collection of raw data

Taking these six principal component factors as input and the total number of EDP as output, a network of six inputs and one output is constructed.

➢ Preprocessing of original data

The following formula is used to format the input and output data of the sample, including

$$\bar{t} = \frac{2(t-t_{min})}{t_{max}-t_{min}} - 1 \quad (6)$$

In the formula: t is the variable before normalization; t_{max} and t_{min} are the maximum and minimum values of t respectively. $\bar{t}$ is a normalized variable.

➢ Network training

It is very convenient to realize the function of artificial neural network by using the neural network toolbox provided by Matlab [4]. Because there are 6 independent variables and 1 dependent variable in the young flow prediction, the number of input neurons is 6, the number of output neurons is 1, and the number of neurons in the middle hidden layer, the BP network needs to be determined according to experience, and the RBF network will be adaptively determined in the process of training.

2.3 Prediction result

Using the matlab toolbox, the total number of people expected to obtain EDP in 2020 and 2021 (unit: ten thousand people) is: 2331, 2617. As shown in the figure below

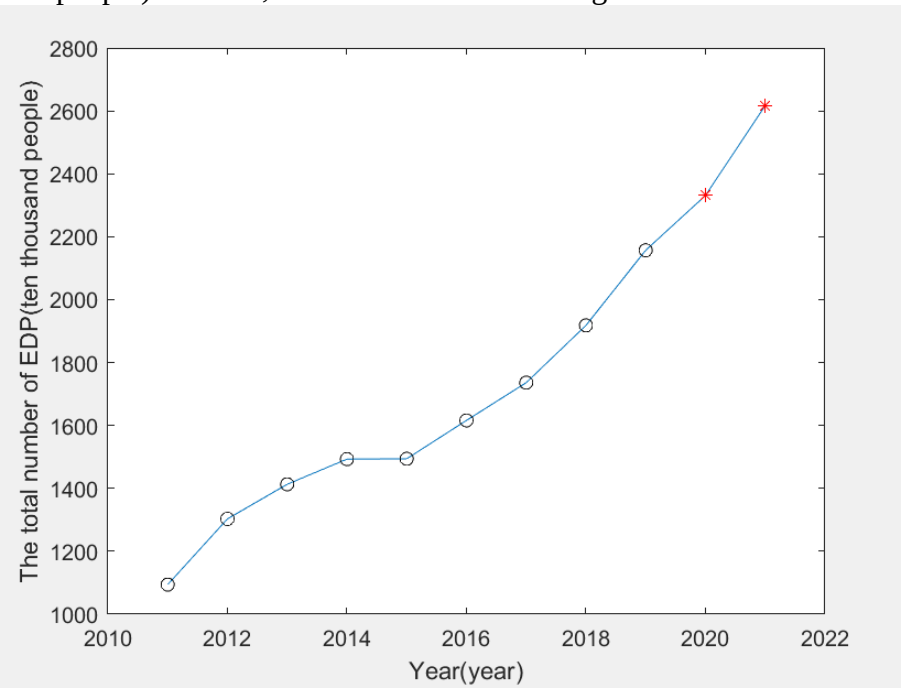

Figure 1: EDP forecast number

3. Risk assessment model [5]

3.1 Step

1. Through the qualitative analysis of the risk of cultural loss, we summarize six risk factors, respectively: the mode of migration, the destination of migration, the religious belief of the migrants and their marital status, physical health, age. For different risk factors, we give different scoring rules, such as:

Migration mode: A_1 is the score coefficient of migration mode, B_1 is the actual travel score of migrants;Walking score is 1, Car score is 2, Ship score is 3, Plane score is 4, Train score is 5.

The sum of a person's scores in the above six items is:

$$X=A_1B_1+A_2B_2+\cdots+A_6B_6 \quad (7)$$

2. According to the materials we collected, two groups were randomly selected from the tables with higher comprehensive scores and those with lower comprehensive scores.

The ratings of the first group are as follows (The second group is the same):

$$X_i = A_1B_{i1} + A_2B_{i2} + \cdots + A_5B_{i5} + A_6B_{i6} \qquad i=1,2,\cdots,6 \tag{8}$$

Order again $\overline{X}^0$is the average of the first group, $\overline{X}^1$is the average of the second group.In order to increase the difference between the two groups, that is, the smaller the sum of squares of the deviation within the group, the better, that is:

$H(A_1, A_2, A_3, A_4, A_5, A_6) = \frac{(\overline{X}^0 - \overline{X}^1)^2}{\sum_{i=1}^{6}(\overline{X_i^0} - \overline{X}^0)^2 + \sum_{i=1}^{6}(\overline{X_i^1} - \overline{X}^1)^2}$, The bigger the better, Thus the score coefficient is the maximum point of (A_i).

The solution of $H(A_i)$can be obtained from the differential equation.

Solution : $(A_1, A_2, A_3, A_4, A_5, A_6)$=(0.9 ,1.6,5.8,4.2,3.8,6.7)

3. Determine the scoring function

$$Y=0.9\,A_1 + 1.6A_2 + 5.8A_3 + 4.2A_4 + 3.8A_5 + 6.7A_6 \tag{9}$$

4. The average of the two groups

$$\overline{X}_A^0 = 83.16$$

$$\overline{X}_B^0 = 37.57$$

Taking the average, the critical value is obtained as follows: 60.365

3.2 Evaluation results

The gradient of cultural risk is divided into:

- 75.583~90.8: the risk of cultural loss is extremely low.
- 60.365~75.583: the risk of cultural loss is low.
- 43.783~60.365: the risk of cultural loss is higher.
- 27.2~43.783: the risk of cultural loss is extremely high.

4. Advantages and disadvantages of the model

1. Advantages

- Based on the risk assessment model, we set different scores for different conditions, which can reasonably divide the cultural loss risk gradient, which is suitable for all kinds of EDP refugees to assess their own cultural loss risk level.
- Based on the principal component analysis model, the workload of index selection can be reduced, and it is easy to use BP neural network to predict the number of EDP at present and in the future.

2. Disadvantages

- Based on the BP neural network, for the case in this paper, when the number of neurons in the hidden layer of the BP network is 6, the calculation results are relatively stable, and when the number of neurons in the hidden layer is taken as other values, the running results are unstable, resulting in errors.
- Considering the situation of EDP refugees only from part of the situation, it is impossible to analyze the ownership of EDP refugees comprehensively and thoroughly.

References

[1] Si Shoukui .Mathematical modeling algorithm and Application [M] National Defense Industry Press, 2015, 231-240
[2] https://www.emdat.be/
[3] Si Shoukui .Mathematical modeling algorithm and Application [M] National Defense Industry Press, 2015, 425-43
[4] Zhuo Jinwu. The application of MATLAB in mathematical modeling [M] Beijing University of Aeronautics and Astronautics Press, 2014-9, 137-159
[5] https://wenku.baidu.com/view/2f6a154bc4da50e2524de518964bcf84b9d52dc0.html