Research on the Ecological Environment Evaluation of Saihanba Based on Entropy Weight Method

Kaibin Xu

College of Mechanical and Electrical Engineering, Sichuan Agricultural University, Ya'an, Sichuan, 625000, China

Abstract: This paper mainly studies the ecological evaluation of Saihanba. First, this paper calculate the Pearson correlation coefficient of the collected Saihanba data indicators, and after obtaining their significant correlations, because too many indicators are not conducive to the calculation, perform cluster analysis on the data, due to the K-means clustering method The subjectivity is too strong, use systematic clustering method to cluster the target, and use the elbow criterion to obtain the K value. Finally import the classified data indicators into MATLAB, and use the entropy weight-TOPSIS model to score and quantitatively evaluate Saihanba's environmental rating.

Keywords: Ecological evaluation, Pearson correlation coefficient, K-means, Entropy weight-TOPSIS model

1. Introduction

Saihanba Forest Farm was originally a vast plateau outside the Great Wall with white sands and sparse trees\(^1\). The species is scarce and the ecosystem is fragile. Moreover, its harsh ecological environment has severely affected the weather conditions in Beijing, which is not conducive to the sustainable development of our country\(^2\).

Since 1962, more and more young people have planted their dreams of restoring forests and repainting a piece of greenery in Saihanba. Today, the ecology of Saihanba has been restored and the water quality has been steadily improved\(^3\). It plays an important role in maintaining ecological balance and stability, and has become an important ecological barrier to guard Beijing and Tianjin. With the new goal of ecological restoration, the people of Saihanba\(^4\) have created three major projects: afforestation, natural improvement of artificial forests, and near-planning and cultivation of natural forests, striving to bring artificial forests closer to natural forests.

2. Model preparation

2.1. Pearson correlation analysis

In statistics, the Pearson product-moment correlation c-efficient is usually represented by r or p, and is used to measure the relationship between two variables X and Y, the value range is between [-1, +1]. The Pearson correlation coefficient is widely used in academic research to measure the strength of the linear correlation between two vari-ables.

First, perform descriptive statistics on Saihanba’s forest coverage, coverage area, forest accumulation, water conservation, carbon dioxide absorption, and oxygen release from 1962 to 2021. Then make a scatter diagram to observe whether there is a linear relationship between each group of variables.

It can be seen from the graph that there is a strong correlation between most of the variables. In the next step, we will calculate the Pearson correlation coefficient and use the corcoef function of MATLAB to find the correlation coefficient matrix between the variables as shown in the following figure 1:

From the data comparison table 1 in the above table, it can be seen that there is a significant correlation between the various data. And the number of indicators is large, and the year is long, the next step is to choose a clustering algorithm to classify the data and indicators, which is convenient for later scoring statistics.
2.2. Cluster analysis

Cluster analysis refers to the analysis process of grouping a collection of physical or abstract objects into multiple classes composed of similar objects. The K-means clustering analysis method in dynamic analysis is used in this paper\(^5\). Since the dimensions of each index are different, first eliminate the dimensions between each index. The results obtained are shown in the Table 2:

<table>
<thead>
<tr>
<th>Table 2: Descriptive Statistics Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Forest cover rate</td>
</tr>
<tr>
<td>Forest coverage area/10,000 mu</td>
</tr>
<tr>
<td>Forest accumulation/10,000 cubic meters</td>
</tr>
<tr>
<td>Conservation of water/10 m³</td>
</tr>
<tr>
<td>CO2 absorption/10,000 tons</td>
</tr>
<tr>
<td>Oxygen release/10,000 tons</td>
</tr>
<tr>
<td>valid cases</td>
</tr>
</tbody>
</table>

After that, perform K-means clustering analysis on the data with dimensions removed, and select the number of clusters to be 2 and 3 respectively for clustering.

When the clustering result is set to 2, the results obtained are based on 1987. The six data indicators after 1987 have been significantly improved compared with those before 1987. The transformed
Saihanba has rich natural resources, improved forest quality, improved water quality, and enhanced ecosystem regulation capabilities.

When the set result is 3, the result obtained distributes the 60-year data at two time points in 1983 and 2001. During 1964-1983, the environment of Saihanba showed a slow growth trend, and in 1983-2001 During the year, the environmental condition of Saihanba was in a transitional period. From 2001 to 2021, the environmental situation of Saihanba was at a rapid growth rate. Optimization, ecological restoration work has made remarkable achievements, and the improvement rate of Saihanba’s ecological environment has shown a rapid growth.

Since the K-means clustering algorithm requires the operator to set the K value, and the system clustering algorithm will be used. Use the elbow criterion to analyze the value of K. The coefficients are stored in the table of data:

$$I(x) = -\ln(p(x))$$, because $0 \leq p(x) \leq 1$, so $I(x) \geq 0$. (1)

The possible situations if event X are $x_1, x_2, \ldots x_n$. Then we can define the information entropy of event X as:

$$H(x) = \sum_{i=1}^{n}[p(x_i)I(x_i)] = -\sum_{i=1}^{n}[p(x_i) \ln(p(x_i))]$$

(2)

It can be seen from the above formula that the essence of information entropy is the expected value of the amount of information.

The basic steps of the entropy method are as follows:

**Step 1:** Build the original matrix

First, according to the data of the past 60 years in Saihanba, the original matrix is constructed:
\[ X = (x_{ij})_{n \times m} \quad (i = 1, 2, 3, ..., n, j = 1, 2, 3, ..., m) \]  

(3)

**Step 2:** Normalized processing data

Normalize the selected indicators to initially eliminate the differences caused by dimensions:

For positive indicators:

\[ Y_{ij} = \frac{x_{ij} - \min(x_{ij})}{\max(x_{ij}) - \min(x_{ij})} \]

For negative indicators:

\[ Y_{ij} = \frac{\max(x_{ij}) - x_{ij}}{\max(x_{ij}) - \min(x_{ij})} \]

**Step 3:** Calculate the entropy

\[ e_j = -\frac{1}{\ln n} \sum_{i=1}^{n} p_{ij} \ln p_{ij} \]  

(4)

**Step 4:** Calculate the index weight

\[ w_i = \frac{1 - e_j}{m - \sum e_j} \quad (j = 1, 2, 3, ..., m) \]  

(5)

**Step 5:** Calculate single index and comprehensive score

The TOPSIS method is more efficient in comprehensive evaluation. Therefore, the entropy weight method and the TOPSIS method are used to process the collected raw data of each province, calculate the weight and convert it into the corresponding score, and then obtain its comprehensive development index $S_j$. Based on this, comprehensive evaluation and analysis of the development of various regions are carried out.

\[ S_{ij} = w_i \times X_{ij} \quad S_i = \sum_j S_{ij} \]  

(6)

4. Model results

Import the data into MATLAB, first standardize the data, and then use the entropy method to determine the weight of each indicator, and get the following Table 3:

<table>
<thead>
<tr>
<th>Forest cover rate</th>
<th>Forest coverage area $/10^4$ mu</th>
<th>Forest accumulation $/10^4$ m$^3$</th>
<th>Conservation of water $/10^4$ m$^3$</th>
<th>CO2 absorption $/10^4$ tons</th>
<th>Oxygen release $/10^4$ tons</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0387</td>
<td>0.0387</td>
<td>0.2363</td>
<td>0.2138</td>
<td>0.2363</td>
<td>0.2363</td>
</tr>
</tbody>
</table>

After that, bring the data divided into three categories in the system cluster into the program to calculate the score and get the score:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0072</td>
<td>0.032</td>
<td>0.053</td>
</tr>
</tbody>
</table>

It can be seen from the score that during 1962-2000, although the ecological environment of Saihanba has been improved, the overall ecological environment is not optimistic. Data show that the forest coverage rate has increased from 13.57% to 69.44%, and the forest coverage area has increased from 190,000 mu to 972,200 mu. In 2000, the forest stock volume reached 5,445,500 cubic meters, the water conservation volume reached 58 million cubic meters, and the carbon dioxide absorption reached 451,800 tons and oxygen release amounted to 314,300 tons. Compared with the data in 1962, these ecological environmental indicators have been improved to varying degrees, but there is still a lot of room for improvement, and the ecosystem is still in a relatively fragile state. Ecological benefits need to be improved, and the regulation capacity of the ecosystem needs to be improved.

During 2001-2019, the ecological environment of Saihanba has been restored to a certain extent, the ecological structure layout has been optimized, and the forest coverage rate has increased to 80.66%. The area has reached 1,129,300 mu, and these two indicators have been steadily improved on the basis of 2000. In addition, the ecological benefits of Saihanba have also been improved. The forest volume has doubled compared with 2000, reaching 10.207 million cubic meters, and the amount of water conserved has tripled compared with 2000, reaching 194 million cubic meters. The amount of carbon dioxide...
absorption has increased. In 2000, it nearly doubled to 846,900 tons, and the amount of oxygen released nearly doubled from 2000 to 589,100 tons. It can be seen that from 2001 to 2019, various ecological environment indicators have been greatly improved, the rate of improvement of the ecological environment is high, and the rate of restoration of the ecological environment of Saihanba has shown a rapid growth trend.

During 2020-2021, Saihanba has changed from barren sandy land to a green pearl, becoming a green miracle of ecological environment management. The forest coverage rate has reached a new high, from 13.57% to 82.21%, an increase of 68.64%; the forest coverage area has attracted worldwide attention, increasing from 190,000 mu to 1.151 million mu, an increase of 961,000 mu in total; Saihanba is rich in natural resources There are 261 species of terrestrial wild vertebrates, 32 species of fish, 660 species of insects, 179 species of macrofungi, and 625 species of plants, which together constitute the rich ecosystem of Saihanba; the ecosystem’s ability to regulate continues The ecological benefits are very optimistic. The forest stock volume ranks in the forefront, as high as 10.368 million cubic meters, which is about 31.42 times the data in 1962; the water quality has been improved, and the amount of conserved water has been continuously increased. It has reached 284 million cubic meters. Compared with 1962, it has increased by about 30.5 times; the carbon dioxide absorption is as high as 860,300 tons, which is about 31.4 times of the 1962 data; the oxygen release has increased from 19,000 tons to 598,400 tons, an increase of nearly 30.5 times. It can be seen that from 1962 to 2021, the improvement of Saihanba's ecological environment has achieved remarkable results, all indicators of the ecological environment have been improved by leaps and bounds, and the ecological benefits have been greatly improved. At present, Saihanba has become an important ecological barrier to guard Beijing and Tianjin, playing a major role in resisting wind and sand, improving the climate and conserving water sources.

5. Conclusion

Through ecological evaluation analysis, the improvement of Saihanba's ecological environment has achieved remarkable results, all indicators of the ecological environment have been improved by leaps and bounds, and the ecological benefits have been greatly improved.

The cluster analysis greatly improves the operating efficiency of the model and facilitates subsequent scoring and evaluation, and the computer is used to assign weights to various indicators to reduce subjectivity. Cluster analysis uses systematic clustering and elbow criteria to determine the K value, which makes the value more accurate and reliable.

References