Ecological Environmental Protection Strategy Based on Big Data Analysis

You Yu 1, 2*
1Department of Environmental Science, Chongqing University, Chongqing, 400044, China
2Chongqing Environmental Protection Information Center, Chongqing 401147, China.
*Corresponding author (Email: youyu808@163.com)

Daijun Zhang
1Department of Environmental Science, Chongqing University, Chongqing, 400044, China
2State Key Laboratory of Coal Mine Disaster Dynamics and Control, Chongqing University, Chongqing, 400044, China

Reza Alavinia
Software College of Northeastern University, Shenyang 110004, China

Jia Deng, Sha Xu
China Environment State Investment(Beijing) Environment Information Technology Co, Ltd. Beijing 100011, China.

Abstract
The issue of ecological environmental protection in China also becomes more and more complex. To better realize the goal of building a resource-saving and environment-friendly society advocated by the government, it is very necessary to apply big data technology in ecological environmental protection. For this end, this paper carries out an in-depth study on the application of big data technology in ecological environmental protection, in the hope of promoting the better unfolding of ecological environmental protection work in China to some extent. To complete the study on the application of big data technology research in ecological environmental protection satisfactorily, we also need to gain a deep understanding of the role of big data technology in ecological environmental protection. In the analysis of big data technology, the author finds that from the perspective of social concern, rapid growth, value relevance and mass data, etc., big data technology can acquire a large number of environmental data and information using feature-rich test equipment under high social concern. Since these data have a high correlation, all kinds of information collected from environmental protection work can be served in big data technology and eventually help better unfold ecological environmental protection.

Key words: Big Data Analysis, Ecological Environmental Protection, Cloud Computing

1. INTRODUCTION
With the rapid development of society and economy in China today, we have made rapid progress in science and technology. This progress enabled China to officially enter the era of big data early. This kind of big data used to achieve the purpose of forecast by analyzing mass data actually change the way Chinese people live and work. In the current ecological environmental protection in China, big data technology can monitor all factors related to ecological environmental protection, such as water, soil, air and organism, etc. comprehensively through remote sensing technology, allowing ecological environmental protection to be better unfolded with the support of big data technology. This is of positive significance for the building an energy-saving and environment-friendly society in China. As early as the 1990s, China began to carry out monitoring work with regard to the issue of ecocological environmental protection and accumulated an enormous number of various types of data related to ecological environment. To realize the evaluation and early warning of all kinds of ecological monitoring data, we must draw support from big data technology. With the support of this technology, the pertinence and effectiveness of ecological environmental protection work will be enhanced.

To complete the study on the application of big data technology research in ecological environmental protection satisfactorily, first we need to gain a deep understanding of big data technology. Fig. 1 shows the global data size forecast. From this figure, we can see that the global transaction data, sensor data, interaction data, data produced by monitoring equipment, etc., have shown an explosive growth trend in recent years. These data are characterized by volume, variety, velocity and value, etc., making big data technology have extremely high application value. As abovementioned, achieving the purpose of forecast by analyzing mass data is the core
of big data technology. This is equivalent to the intelligent Internet, which provides us with specific service. In this kind of service, big data technology can predict the possibility of relevant events, based on mass data. This brings a lot of promoting effect for the better unfolding of ecological environmental protection work (Bigano and Proost, 2003).

![Global Data Size Forecast](image)

**Figure 1. Global Data Size Forecast**

2. **BIG DATA Analysis IN ECOLOGICAL ENVIRONMENTAL PROTECTION**

To complete the study on the application of big data technology research in ecological environmental protection satisfactorily, we also need to gain a deep understanding of the role of big data technology in ecological environmental protection.

In the analysis of big data technology, the author finds that from the perspective of social concern, rapid growth, value relevance and mass data, etc., big data technology can acquire a large number of environmental data and information using feature-rich test equipment under high social concern. Since these data have a high correlation, all kinds of information collected from environmental protection work can be served in big data technology and eventually help better unfold ecological environmental protection in China. Combined with relevant literature and my own knowledge, the author summarizes the role of big data technology in ecological environmental protection into three aspects: to identify a development goal, to control the overall layout and to build a big data application center (Costapierce, 2010).

2.1. **Identify a Development Goal**

As far as ecological environmental protection work is concerned, the use of big data technology can better identify a development goal for ecological environmental protection work, which has a good guiding effect on ecological environmental protection work.

With the rapid development of economy and society in China today, both Chinese people and society attach great importance to ecological environmental protection work. Environmental monitoring by an environmental protection agency and administrative reforms by a supervision and law enforcement agency also greatly improves the effectiveness of this work. However, at present, major environmental protection agencies in China lack consistency in data collection, sorting, summary and analysis. This has brought a considerable negative effect on the better unfolding of environmental monitoring, environmental impact assessment and environmental law enforcement, etc. While the application of big data technology can better realize the integrated management and analysis of this kind of data, which definitely makes all kinds of ecological environmental protection work in China better complete their functions under the guidance of a clear goal (Dearfield, Bender and Kravitz, 2009).

2.2. **Control the Overall Layout**

To guarantee that big data related to ecological protection could be better applied to ecological environmental protection work, big data technology must also consider ecological environmental protection work and achieve the overall layout of big data related to ecological environmental protection, to ensure that big data related to ecological environmental protection play a proper role and truly support the better unfolding of environmental protection work in China (Dearfield, Bender and Kravitz, 2009).

Regional compound ecological system is an ecological function entity jointly composed of human society, economic activities and natural conditions. In the society-economy-nature regional compound ecological system, three subsystems are mutually correlative and restrictive to form a complicated ecological relationship. If \( M_1 \), \( M_2 \) and \( M_3 \) refer to three subsystems respectively; \( J_i \), \( Y_i \) and \( G_i \) refer to the structure, elements and functions...
of the subsystem $M_i$; $R_j$ refers to the incidence relation among the subsystems within the system; $N$ refers to research object; and $T$ refers to time. Therefore, regional compound ecological system can be expressed as the following model:

$$S \subseteq \{M_1, M_2, M_3, R, T, N\} \quad S \subseteq \{J, Y, G\} \quad (i = 1, 2, 3)$$

2.3. Build a Big Data Application Center

As for ecological environmental protection work in China, to maximize the role of big data technology, Chinese government must identify an environmental information resources acquisition standard. With this standard, big data technology can be applied to establish an ecological environment monitoring and data acquisition network and advanced environmental monitoring early warning and assessment system, so that big data technology can better support the ecological environmental protection work in China (Fargione, Plevin and Hill, 2010).

3. PRACTICE OF BIG DATA ANALYSIS IN ECOLOGICAL ENVIRONMENTAL PROTECTION

For ecological environmental protection work that big data technology is applied to, whether this work can be better unfolded depends “30% on technology and 70% on data”. This shows the importance of data for this application. According to the author’s investigation, it is found that an environmental quality monitoring system based on big data technology has been successfully applied to many provinces in China at present. To gain a deeper understanding of the existing practice of this big data technology in ecological environmental protection, below the author will elaborate on the design and implementation of this monitoring system.

3.1. The Design of an Environmental Quality Monitoring System Based on Big Data Technology

Fig. 2 shows a topological structure of environmental quality monitoring business. From this figure, we can find the overall workflow of this environmental quality monitoring system based on big data technology. However, since RS, GIS, IOT and other advanced environmental technology are all applied to this system, the original environmental quality monitoring system cannot well satisfy the application of big data technology. So it seems very necessary to design an environmental quality monitoring system based on big data technology (Gallardo-Vázquez, Sánchez-Hernández, 2014).

![Figure 2. The Topological Structure of Environmental Quality Monitoring Business](image-url)
Table 1. Summary of System Demands

<table>
<thead>
<tr>
<th>System Demands</th>
<th>Specific Demands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Communication</td>
<td>To collect monitoring data and store them in a data communication server. Store merged files in the data storage layer of system.</td>
</tr>
<tr>
<td>Data Storage</td>
<td>Store PH, chemical oxygen demand (COD), NO2, PM2.5 and other data that are collected in a database. The database must have high stability, high scalability, high reliability, high availability and mass data storage.</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>To query environmental quality monitoring data, analyze environmental quality monitoring data statistically, reflect environmental quality conditions, give early warnings on abnormal environments, analyze the law of environmental quality monitoring data and forecast the variation trend of environmental quality</td>
</tr>
<tr>
<td>Data Application</td>
<td>To display results using simple tables or figures and ensure the access quality</td>
</tr>
</tbody>
</table>

Figure 3. Logic Framework of the System

Fig. 4 shows the logic framework of the environmental quality monitoring system based on big data technology. From this figure, we can find that the framework design of this system is divided into six parts: data communication layer, source data storage layer, data preprocessing layer, hybrid storage layer, data access layer and data application layer.

Figure 4. The Framework of the Environmental Quality Monitoring System Based on Big Data Technology

The design of the data communication layer of framework of environmental quality monitoring system based on big data technology is mainly divided into five modules, i.e., data file load module, data file merge module, data file upload module, data communication layer task scheduling module and configuration information management module. Tab. 2 shows the functions of these five modules in detail.

Table 2. The Functions of Data Communication Layer Modules

<table>
<thead>
<tr>
<th>Module</th>
<th>Functions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data File Load Module</td>
<td>To load an environmental quality monitoring data file and store it in a data communication server.</td>
</tr>
<tr>
<td>Data File Merge Module</td>
<td>To merge environmental quality monitoring data files according to merge rules.</td>
</tr>
<tr>
<td>Data File Upload Module</td>
<td>To upload an environmental quality monitoring data file to a Hadoop cluster.</td>
</tr>
<tr>
<td>Data Communication Layer Task Scheduling Module</td>
<td>To perform data acquisition according to time point and time interval</td>
</tr>
<tr>
<td>Configuration Information Management Module</td>
<td>To store and manage all kinds of configuration information</td>
</tr>
</tbody>
</table>
In the design of source data storage layer, this layer is mainly responsible for storing and managing the original environmental quality monitoring data. The reading and writing of environmental quality monitoring data files are two major jobs. In the design of source data preprocessing layer, as this layer needs to complete the data noise data processing of environmental quality monitoring data, it is mainly divided into two parts: data preprocessing based on Map Reduce and data preprocessing based on Hive. While in the design of hybrid storage layer, this layer can improve the expansibility of monitoring data and data storage capacity and better satisfy the application of analysis algorithm. Data storage design and data import based on HBase is the key to the design of this layer. Fig. 5 shows the flowchart of data import in HBase visually (Kobing, Raburu and Masese, 2009).

**Figure 5.** Flowchart of Data Import in Hbase

In the design of data access layer, it can be divided into HBase-based data access and Oracle-based data access. In the design of data application layer, it needs to satisfy the query, evaluation, trend prediction and other demands of environmental quality monitoring data. Tab. 3 shows these demands in detail.

**Table 3.** The Functions of Data Access Layer

<table>
<thead>
<tr>
<th>Demand of Data Access Layer</th>
<th>Specific Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Query</td>
<td>To query environmental quality monitoring data</td>
</tr>
<tr>
<td>Data Evaluation</td>
<td>To standardize monitoring data and calculate and obtain the evaluation level</td>
</tr>
<tr>
<td>Trend Prediction</td>
<td>To draw a line chart automatically to visualize changes in environmental quality monitoring factors</td>
</tr>
<tr>
<td>Over-range Alarm</td>
<td>To show whether environmental quality monitoring data exceeds the monitoring range</td>
</tr>
<tr>
<td>Data Simulation</td>
<td>To understand the possible environmental impacts of over-range of specific indicators</td>
</tr>
</tbody>
</table>

### 3.2. Forewarning Cloud Model of Ecological Safety

Until now, there are two main categories of ecological safety forewarning cloud models in China:

Direct application of ecological safety and quality index to present ecological status quality and future variation tendency. According to the ecological quality of objects and indicators of environmental quality, quantize with continuous real number interval \([0,10]\) to express the quality variation scope. When the eco-environmental quality is under the most ideal state, the value is 10. And the value is 0 under the worst state.

The development trend of ecological security of ecological function area is reflected according to the warning conditions reflected by utilizing certain calculation method through the early-warning index system of components. The commonly used classification is divided to two types:

The first type is to divide ecological security early-warning to bad condition forewarning, negative evolution forewarning and deterioration speed forewarning. It is assumed that at \(t\), the ecology and environment can be subject to forewarning in the coming certain period, and the forewarning period is
\( \Delta t = T - t \); \( E(t) \) refers to ecological environment quality rating value; parameter \( EP \) expresses the critical value of ecological and environment quality average value of defective state warning; parameter \( \Delta EP \) and \( \Delta EP_2 \) refer to deteriorating trend warning and deterioration speed warning respectively; is the critical value of quality average value at the period \( \Delta t \); if the forewarning evaluation object is random with uncertainty, namely assurance rate \( a \) parameters are as the constraints for forewarning evaluation. Therefore, under the given guarantee rate \( a \) and the parameters of \( EP, \Delta EP \) and \( \Delta EP_2 \), the ecological environment and the mathematical model of safety forewarning evaluation and determination are:

Defective status forewarning: \( W \{ EP_1 < E(t) < EP_2 \} > a \)

Whereas in the formula \( EP_1 = 2, EP_2 = 4 \) refers to poor state forewarning; \( EP_1 = 0 \) and \( EP_2 = 2 \) refer to defective state forewarning.

Deterioration tendency forewarning: \( W \{ E(T) < E(t) < EP_1, E(T) - E(t) \geq \Delta EP \} \geq a \)

Deterioration speed forewarning: \( W \{ E(T) < E(t) < EP_1, E(T) - E(t) > \Delta EP_2 \} \geq a \)

Whereas in the formula, \( \Delta EP \) can be determined according to limited time period of river valley reclamation project, such as at 5-year planning \( \Delta EP = 1/5(1/a) \)

Generally, it can be determined by adopting target warning coefficient. The formula is:

\[
R = \frac{X_i}{X_0} - 1
\]  \hspace{1cm} (1)

Whereas in the formula, \( R \) -Forewarning system;

\( X_i \) -Actual value of forewarning index;

\( X_0 \) -Early warning target value.

\( R \) It can be divided into the following three circumstances according to the nature of the warning indicators:

\( R > 0 \) Warning; \( R < 0 \) No warning; \( R = 0 \) Intermediate warning.

After confirming warning information according to the above model, a series of early-warning processes will be involved until reaching the final warning degree.

The other type is: To define forewarn interval and determine the boundaries of forewarning type. Generally, red, orange, yellow, green and blue are used to display five kinds of ecological security statuses for ecological safety as for classification. The forewarning cloud model based on matter element analysis is introduced right here to determine the overall index of warning alarm, and to determine the ecological security status. Professor Caiven, a Chinese scholar, proposed matter element analysis theory, established a matter element model \( R = (N, C, V) \). \( N \) is the name of object. \( C \) is the characteristic of object \( N \). And \( V \) is the measurement value of the characteristic \( C \) of the object \( N \). Ternary ordered group is named as the basic element of object. Then as described above, forewarning is divided into five ecological warning levels \( N_1, N_2, N_3, N_4, N_5 \).

Corresponding matter-element is summarized and established as:

\[
R_j = \left[ \begin{array}{c}
N_j, C_1, V_1 \\
C_2, V_2 \\
\vdots \\
C_m, V_m
\end{array} \right] = \left[ \begin{array}{c}
N_j, C_1, \left\{ a_{ji}, b_{ji} \right\} \\
C_2, \left\{ a_{j2}, b_{j2} \right\} \\
\vdots \\
C_m, \left\{ a_{jm}, b_{jm} \right\}
\end{array} \right]
\]  \hspace{1cm} (2)

\( N_j \) refers to \( j \) security levels. \( V_\mu \) is the value range \( \left\{ a_{\mu}, b_{\mu} \right\} \) of the characteristic \( C \) of \( N_j \). \( R_j \) is the classic language. The sectional field of corresponding classical language is \( R_p \) as below. In addition \( R_p \supset R_j \):

\[
R_p = \left[ \begin{array}{c}
N_p, C_1, V_1 \\
C_2, V_2 \\
\vdots \\
C_m, V_m
\end{array} \right] = \left[ \begin{array}{c}
N_p, C_1, \left\{ a_{p1}, b_{p1} \right\} \\
C_2, \left\{ a_{p2}, b_{p2} \right\} \\
\vdots \\
C_m, \left\{ a_{pm}, b_{pm} \right\}
\end{array} \right]
\]  \hspace{1cm} (3)
Where, \( N_p \) refers to the overall range scope. Similarly, \( V_p \) is the value range \( \{a_p, b_p\} \) of the characteristic \( C_i \) of \( N_p \). The matter-element defining forewarning object is expressed as:

\[
R_s = (p_s, c_s, v_s)
\]

The status value \( v_i \) of the forewarning index of forewarning object belongs to certain safety level degree which is equivalent to membership in fuzzy mathematics. The difference is that the membership scope is zone [0,1], which will be reflected to real number axis \([-\infty, +\infty]\). The value of \( i \) index defining forewarning object associated with the \( j \) safety level scope is:

\[
F_j (v_i) = \begin{cases} 
\rho(v_i, v_{p_j}) \rho(v_i, v_{p_j}) - \rho(v_{p_j}, v_{p_j}) - \rho(v_{p_j}, v_{p_j}) \neq 0 \\
-\rho(v_{p_j}, v_{p_j}) - 1, \rho(v_i, v_{p_j}) - \rho(v_{p_j}, v_{p_j}) = 0
\end{cases}
\]

(5)

Where, \( v_{p_j} \) is the scope of classic language; \( V_{p_j} \) is the scope of sectional field; \( \rho(v_i, v_{p_j}) \) and \( \rho(v_{p_j}, v_{p_j}) \) are the distance of \( \{a, b\} \) and \( \{a, b\} \) within the scope range of the value \( c_i \) and \( v_i \).

\[
\rho(v_i, v_{p_j}) = \frac{v_i - a_p + b_p}{2} - \frac{a_p - b_p}{2}; \quad \rho(v_i, v_{p_j}) = \frac{v_i - a_p + b_p}{2} - \frac{a_p - b_p}{2}
\]

(6)

Take \( G_i (v_i) = \max F_j (v_i) \) and \( j \in \{1, 2, \cdots, m\} \). Forewarning index \( v_i \) belongs to level \( j \). The general index status of forewarning object is available \( F_j (R_s) = \sum w_i F_j (v_i) \). \( w_i \) is the weighting coefficient of index \( i \) which can be determined by entropy weight method etc.. Then in case of \( G_j (R_s) = \max F_j (R_s) \) and \( j \in \{1, 2, \cdots, m\} \), the overall forewarning object \( R_s \) belongs to level \( j \). Eventually, forewarning object shall make red, orange, yellow, green and blue forewarning according to \( j \), the registration range.

### 3.3. The Implementation of Environmental Quality Monitoring System Based on Big Data Technology

In the implementation of environmental quality monitoring system based on big data technology, we need to combine the abovementioned environmental quality monitoring system framework based on big data technology. The implementation of six parts of framework design will be elaborated as follows.

The implementation of data communication layer must be done on the basis of Quartz task scheduling framework. To guarantee that data communication layer can manage and access configuration information quickly, we store configuration information using figures and charts. Tab. 4 shows part of this configuration information. Besides, in the implementation of data communication layer, we also need to apply a time configuration table to clarify the timing for scheduling. Due to the limited length, this paper will not dwell on this time configuration table. (Osuna, Börner and Nehren, 2014).

<table>
<thead>
<tr>
<th>Field Name</th>
<th>Data Type</th>
<th>Field Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Monitoring URL</td>
<td>nVarchar (255)</td>
<td>IP Address</td>
</tr>
<tr>
<td>longitude</td>
<td>nVarchar (255)</td>
<td>Longitude</td>
</tr>
<tr>
<td>latitude</td>
<td>nVarchar (255)</td>
<td>Latitude</td>
</tr>
</tbody>
</table>

In the implementation of source data file storage layer, File Shell, File System API, Web HDFS can better satisfy the demands of source data file storage layer. In the implementation of data preprocessing layer, both Map Reduce programming model and Hive data warehouse tool can satisfy the demands of data preprocessing layer. Both implementations require identifying the data structure of environmental quality monitoring data records. Tab. 5 describes part of this structure in detail.
randomness, entropy, the worse condensation of quantitative value represents condensation degree of qualitative linguistic value certainty. The higher hyper degree of be wider, and the concept will be more ambiguous; understood as ambiguity. The higher expected value similarity degree measurement which reflects the acceptable range of qualitative linguistic value which can to the coordinate system of the qualitative concept; it is the center variables: Expected Value

4.

such an environmental quality monitoring system based on big data technology has achieved good applications in many provinces in China. Combined with big data technology, this system is capable of realizing the storage and analysis of mass environmental data. It allows the laydown of related policies and regulations and better unfolding of ecological environmental protection work to be better supported and illustrates that big data technology can provide better services for ecological environmental protection work in China.

4. SIMULATION ON THE EARLY WARNING CLOUD MODEL OF ECOLOGICAL SECURITY

The digital characteristics of early warning cloud model of ecological security are expressed with three variables: Expected Value $E_x$, Entropy $E_n$ and Hyper Entropy $H_e$. Among them, $E_x$ can accurately reflect the qualitative concept. It is the center-of-gravity position of ecological cloud droplet, namely 100% belonging to the coordinate system of the qualitative concept; $E_n$ refers to fuzziness of qualitative concept, namely similarity degree measurement which reflects the acceptable range of qualitative linguistic value which can be understood as ambiguity. The higher expected value $E_x$, the acceptable value range of qualitative concept will be wider, and the concept will be more ambiguous; $H_e$ is the entropy of entropy $E_n$, namely the dispersion degree of $E_n$, which reflects the condensed dispersion degree of ecological cloud droplet. Actually, each quantitative value represents condensation degree of qualitative linguistic value certainty. The higher hyper entropy, the worse condensation of ecological cloud droplet will be. The higher dispersion, the higher randomness of qualitative concept membership degree will be and the thicker of ecological cloud droplet will be.

<table>
<thead>
<tr>
<th>Monitoring points</th>
<th>Monitoring time</th>
<th>copper (mg/l)</th>
<th>total nitrogen (mg/l)</th>
<th>total phosphorus (mg/l)</th>
<th>Helium and nitrogen (mg/l)</th>
<th>Five day biochemical oxygen demand (mg/l)</th>
<th>Chemical oxygen demand (mg/l)</th>
<th>Permanganate index (mg/l)</th>
<th>dissolved oxygen (mg/l)</th>
<th>PH value</th>
<th>water temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real reservoir</td>
<td>201600710206000</td>
<td>0.674</td>
<td>1.381</td>
<td>0.005</td>
<td>0.090</td>
<td>0.175</td>
<td>0.260</td>
<td>2.350</td>
<td>7.180</td>
<td>7.63</td>
<td>22.1</td>
</tr>
<tr>
<td>Real reservoir</td>
<td>201600221206000</td>
<td>0.700</td>
<td>2.179</td>
<td>0.011</td>
<td>0.060</td>
<td>0.109</td>
<td>0.158</td>
<td>1.350</td>
<td>10.390</td>
<td>7.69</td>
<td>23.2</td>
</tr>
<tr>
<td>Real reservoir</td>
<td>201600241206000</td>
<td>0.716</td>
<td>2.207</td>
<td>0.009</td>
<td>0.020</td>
<td>0.031</td>
<td>0.042</td>
<td>1.380</td>
<td>9.930</td>
<td>7.75</td>
<td>22.1</td>
</tr>
<tr>
<td>Gaube Bridge</td>
<td>201600250000000</td>
<td>0.739</td>
<td>2.748</td>
<td>0.012</td>
<td>0.050</td>
<td>0.088</td>
<td>0.125</td>
<td>4.270</td>
<td>6.840</td>
<td>7.81</td>
<td>23.4</td>
</tr>
<tr>
<td>Gaube Bridge</td>
<td>201600191206000</td>
<td>0.760</td>
<td>3.161</td>
<td>0.014</td>
<td>0.080</td>
<td>0.146</td>
<td>0.211</td>
<td>4.690</td>
<td>6.220</td>
<td>7.77</td>
<td>22.5</td>
</tr>
<tr>
<td>The Yalu River</td>
<td>201600212060000</td>
<td>0.781</td>
<td>3.574</td>
<td>0.016</td>
<td>0.100</td>
<td>0.184</td>
<td>0.267</td>
<td>2.300</td>
<td>9.490</td>
<td>7.76</td>
<td>20.2</td>
</tr>
<tr>
<td>The Yalu River</td>
<td>201600181206000</td>
<td>0.802</td>
<td>3.987</td>
<td>0.018</td>
<td>0.260</td>
<td>0.502</td>
<td>0.743</td>
<td>2.800</td>
<td>9.200</td>
<td>7.83</td>
<td>19.3</td>
</tr>
<tr>
<td>Tang River Reservoir</td>
<td>201600401206000</td>
<td>0.823</td>
<td>4.400</td>
<td>0.020</td>
<td>0.020</td>
<td>0.020</td>
<td>0.019</td>
<td>1.900</td>
<td>6.510</td>
<td>7.86</td>
<td>18.3</td>
</tr>
<tr>
<td>Tang River Reservoir</td>
<td>201600512060000</td>
<td>0.844</td>
<td>4.813</td>
<td>0.022</td>
<td>0.020</td>
<td>0.018</td>
<td>0.015</td>
<td>1.900</td>
<td>6.500</td>
<td>7.82</td>
<td>19.9</td>
</tr>
<tr>
<td>Tang River Reservoir</td>
<td>201600512060000</td>
<td>0.865</td>
<td>5.226</td>
<td>0.024</td>
<td>0.010</td>
<td>-0.004</td>
<td>-0.019</td>
<td>2.100</td>
<td>6.670</td>
<td>7.83</td>
<td>22.5</td>
</tr>
</tbody>
</table>
Actual demands and applications have obtained universality of normal distribution. In addition, normal cloud has been widely applied. The mathematical expectation curve \( MEC \) is expressed in formula 7:

\[
MEC(x) = \exp\left(-\frac{(x - Ex)^2}{2En^2}\right)
\]  

(7)

The normal early warning cloud model of ecological security of linguistic concept “about 20 people” is as shown in Figure 6, where \( Ex = 20 \), \( En = 0.1 \), \( He = 0.02 \) and \( N = 400 \). According to the principle of early warning cloud model of ecological security \( 3En \), the elements beyond the model \([Ex - 3En, Ex + 3En]\) can be ignored.

![Figure 6. Phase aberration Ecological security early warning cloud model](image)

4.1. Ecological Security Early Warning Cloud Generator

4.1.1 Simulation on Forward Ecological Cloud Generator

Forward Cloud Generator is the mapping from the qualitative to quantitative. It expresses ecological cloud droplet according to digital characteristics of early warning cloud model of ecological security (Ex, En and He), as shown in Figure 7.

Specific steps:

Step 1: Generate a normal random number \( X \) with the expected number \( Ex \) and the standard deviation \( En \) (The generation of normal distributed random numbers is subject to identified and mature method);

Step 2: Generate a normal random number \( En' \) with the expected number \( En \) and the standard deviation \( He \);

Step 3: \( x \) is defined as a specific quantitative value of qualitative concept \( A \), which is called as ecological cloud droplet;

Step 4: Substitute the values obtained into formula 8 for calculation:

\[
y = \exp\left(-\frac{(x - Ex)^2}{2(En')^2}\right)
\]  

(8)

Step 5, \( y \) is called as \( x \) the membership of qualitative concept \( A \);

Step 6, then the group of data \((x, y)\) can accurately reflect qualitative and quantitative transformation;

Step 7, repeatedly execute step 1 to step 6 until completely completing the ecological cloud droplet \( N \).
Figure 7. The results of different Initial ecological security early warning based on Forward Ecological Cloud Generator

4.1.2 Simulation on Backward Ecological Cloud Generator

Backward ecological cloud generator is a transformation model realizing transforming from quantitative value to qualitative concept. It can transform certain number of accurate data to qualitative concept represented by digital features (Ex, En and He), as shown in Figure 8.

Figure 8. The results of different Initial ecological security early warning based on Backward Ecological Cloud Generator

Specific steps:
Step 1: Take the value of formula 8 as the estimated value of $\hat{E}_X$:

$$\hat{E}_X = \frac{1}{n} \sum_{i=1}^{n} x_i \quad (9)$$

Step 2: Keep $m$ ecological cloud droplet and delete the remaining points $y > 0.999$;

Step 3: Derive $(x, y')$ according to formula 9 based on $E_x$ and $(x, y)$:

$$E_{n'} = \left| x - \hat{E}_x \right| \left( \frac{1}{\sqrt{2 \ln y}} \right) \quad (10)$$

Step 4: Derive the estimated value $\hat{E}_n$ of $E_n$ according to formula 10;

$$\hat{E}_n = \frac{1}{m} \sum_{i=1}^{m} E_{n_i} \quad (11)$$

Step 5: Derive the estimated value $\hat{H}_e$ of $H_e$ according to formula 11.
Therefore, $\hat{N}$ ecological cloud droplet expresses three characteristic values of qualitative concept, thus to clarify the qualitative concept according to the actual situation.

4.2. Fuzzy Data Mining and Analysis Based on Early Warning Cloud Model of Ecological Security

Perform knowledge mining by adopting early warning cloud model of ecological security. Generally, first of all dig out the several qualitative concepts of objects, divide according to each qualitative concept on fuzzy degree, for example the academic performance can be divided to excellent, good, middle and poor. And then establish an early warning cloud model of ecological security, determine the digital characteristics of early warning cloud model of ecological security, eventually synthesize each concept, and excavate valuable information based on comprehensive fuzzy sets and related indicators.

The concept or knowledge can be divided to $m$ types $(a_1,a_2,\cdots,a_m)$ according to the theory and reality of specific field. Each type represents a valuable classification. Exact $n$ characterization factors $(x_1,x_2,\cdots,x_n)$. Each characterization factor has corresponding actual contents (Multiple fuzzy partition can be contained) and each implication is corresponding to a value (Can be a numerical interval).

According to previous characterization factors extracted, corresponding ownership type fuzzy sets shall be defined according to practical problem $(A_1,A_2,\cdots,A_n)$. Establishment of membership early warning cloud model of ecological security. Determine the early warning cloud $n$ of fuzzy set $(A_1,A_2,\cdots,A_n)$, namely the three digital features $(E_x,E_n,H_e)$ of $n$ early warning cloud sets. Three digital features of $n$ early warning cloud of fuzzy set according to the statistical analysis and calculation respectively:

$$A_i(E_x,E_n,H_e), A_i(E_x,E_n,H_e),\cdots, A_i(E_x,E_n,H_e).$$

Calculate the membership of each feature factor relative to fuzzy set according to three digit characters by utilizing 4.3.1 forward cloud generator algorithm $\mu_{\lambda_i}(x_{i=1,2,\cdots,n})$.

Set $f(x)=(x_1,x_2,\cdots,x_n)$, namely $f(x)$ refers to specific location model of knowledge. And then make the comprehensive fuzzy set $H=A_1 \oplus A_2 \oplus \cdots \oplus A_n$ to indicate a comprehensive horizontal indicator which is defined as:

$$\mu_{\lambda_i}(x)=\begin{cases} 1 & \mu_{\lambda_i}(x)=1, i=1,2,\cdots,n \\ \sum \partial_i \mu_{\lambda_i}(x) & \mu_{\lambda_i}(x) \neq 1, i=1,2,\cdots,n \end{cases}$$

Where, $\partial_i (i=1,2,\cdots,n)$ refers to weight which can be determined according to the simulation data and specific situation, and $\sum \partial_i =1, (i=1,2,\cdots,n)$.

Classify the mined information according to the concept or knowledge defined in 4.1:

$$\begin{align*}
\text{Class } a_1 & : \mu(x) \geq \lambda_i \\
\text{Class } a_2 & : \lambda_i \leq \mu(x) < \lambda_i \\
\vdots & \\
\text{Class } a_m & : \mu(x) < \lambda_{m-1}
\end{align*}$$

Where, $\lambda_i, \lambda_2,\cdots,\lambda_{m-1}$ is the parameter index after analysis and selection. Since the concept classification is ambiguous, the class information and selection of $\lambda$ are determined according to the demands and actual situation.

4.3 Effectiveness Verification of the Model

Perform effectiveness verification for the model to verify if the running results of the model are in line with the actual situation, and if the model is correct and effective. The simulated practice of model inspection is performed from 2006 to 2010. Perform model effectiveness verification specific to 8 representative indicators. Review relevant statistical almanac and information of the areas covered by Qilian glacier and water conservation ecological function. Perform simulation for historical data by adopting running model. Compare the obtained resulting data and actual data. If the error is less than 10%, the reliability of the model can be
verified within reasonable error scope. The comparison of the simulated value and actual value of the four
extracted indicators from 2008–2010 are as shown in Table 7.

**Table 7. Effectiveness verification results of SD model**

<table>
<thead>
<tr>
<th>Year</th>
<th>Per capita disposable income (ten thousand yuan)</th>
<th>Nonagricultural population (ten thousand yuan)</th>
<th>agricultural acreage (Thousands of hectares)</th>
<th>Discharge amount of three industrial wastes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value of simulation</td>
<td>Actual value</td>
<td>Relative error</td>
<td>Value of simulation</td>
</tr>
<tr>
<td>2008</td>
<td>0.208</td>
<td>0.214</td>
<td>0.006</td>
<td>226.51</td>
</tr>
<tr>
<td>2009</td>
<td>0.235</td>
<td>0.227</td>
<td>0.008</td>
<td>253.62</td>
</tr>
<tr>
<td>2010</td>
<td>0.276</td>
<td>0.268</td>
<td>0.008</td>
<td>286.37</td>
</tr>
<tr>
<td>2011</td>
<td>0.308</td>
<td>0.290</td>
<td>0.009</td>
<td>315.36</td>
</tr>
<tr>
<td>2012</td>
<td>0.342</td>
<td>0.317</td>
<td>0.010</td>
<td>345.29</td>
</tr>
<tr>
<td>2013</td>
<td>0.376</td>
<td>0.344</td>
<td>0.011</td>
<td>375.22</td>
</tr>
</tbody>
</table>

**Figure 9.** Trend of effectiveness verification results of ecological security early warning cloud model
Through the data as shown in the table, the error data shall be no greater than 10%. The model obtained can reflect the actual situation and verify the effectiveness of the model. Perform model simulation on this basis. The influence source of the system, accumulation process, space accumulation, time accumulation, as well as the changes of structure and functions can be analyzed via the model simulation. The most urgent issue is to obtain the data and changing tendency required in the forewarning of the next chapter thus to provide effective data and model support for subsequent analysis and warning categories.

5. SUGGESTIONS ON THE BETTER USE OF BIG DATA TECHNOLOGY IN ECOLOGICAL ENVIRONMENTAL PROTECTION

Combined with the above content, we have gained a detailed understanding of big data technology, the role of big data technology in ecological environmental protection and existing practice of big data technology in ecological environmental protection. Hereinafter, the author will combine relevant literatures with my own survey, come up with suggestions on the better use of big data technology in ecological environmental protection. Hopefully, these suggestions would be of certain help for the better unfolding of ecological environmental protection in China.

5.1. To Implement Big Data Engineering Projects Related to Ecological Environmental Protection

Big projects can drive big progress. Thus, to ensure that big data technology can be better applied to ecological environmental protection work, the Chinese government must take the lead to carry out big data engineering projects related to ecological environmental protection, to fan out from a point to an area and gradually expand the application scope of big data technology in ecological environmental protection. Achieve the building of environmental protection big data industry, by integrating environmental protection system data, promoting the sharing and openness of data related to environmental protection, exploring a new environmental monitoring mode, etc., so that the ecological environmental protection work in China can be better unfolded (Rehdanz and Maddison, 2008).

5.2. To Consolidate the Application Foundation

To ensure that big data technology can better serve ecological environmental protection, the Chinese government must also consolidate the application foundation of big data technology. Tab. 8 shows exiting problems with environmental protection department data in detail. From this table, we may be aware of the important significance to improve the authenticity, credibility, availability, usefulness and practicality of environmental protection department data in China.

<table>
<thead>
<tr>
<th>Problems</th>
<th>Influence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low data credibility</td>
<td>Widespread criticism from the society</td>
</tr>
<tr>
<td>Low degree of openness and sharing</td>
<td>Better application of big data technology is affected</td>
</tr>
<tr>
<td>Forged data</td>
<td>Errors in data analysis results</td>
</tr>
<tr>
<td>Inaccurate data</td>
<td>Errors in data analysis structure</td>
</tr>
</tbody>
</table>

5.3. To Promote Demonstrative Applications

Since the applications of big data technology itself are still in its infancy in the ecological environmental protection in China, to give full play to their functions, we need to realize the demonstrative applications of big data ecological technology in various regions across China. Environmental monitoring, environmental impact assessment, environmental law enforcement and environmental risk and so on are all good choices for these demonstrative applications (Rong, Chen and Wang, 2014).

To ensure that big data technology can be better applied to ecological environmental protection work in China, the Chinese government must also perfect supporting policies as soon as possible, to provide legal support for the application of this technology. Meanwhile, increase financial support in this regard. In doing so, the living environment work in China can be better effected.

6. CONCLUSION

In the study on the application of big data technology in ecological environmental protection in this paper, the author elaborates on the concept of big data technology, the role of big data technology in ecological environmental protection, existing practice of big data technology in ecological environmental protection and suggestions on the better use of big data technology in ecological environmental protection. Regional compound ecological system is an ecological function entity jointly composed of human society, economic activities and natural conditions. In the society-economy-nature regional compound ecological system, three subsystems are
mutually correlative and restrictive to form a complicated ecological relationship. The most urgent issue is to obtain the data and changing tendency required in the forewarning of the next chapter thus to provide effective data and model support for subsequent analysis and warning categories. With this series of discussions, we also find big data technology itself can be better applied to ecological environmental protection work in China. To ensure that this application plays a better role, the Chinese government must be highly supportive of this application.

REFERENCES: