Investigation on Data Mining of Intelligent Environmental Protection Big Data Based on Neural Networks

Chao Zhou[1](✉), Junwen Deng[2], and Haoxuan Tang[2]

[1] Powerchina Guiyang Engineering Corporation Limited, Guiyang 550081, Guizhou, China
Zc125@126.com

[2] Guoneng Qinghai Yellow River Maerdang Hydropower Development Co., Ltd., Guoluo 814099, Qinghai, China

Abstract. Intelligent environmental protection refers to the intelligent form of utilizing environmental data information, including comprehensive collection and analysis, continued development and utilization of environmental data, and achieving the goal of supporting environmental management. It is conducive to reducing pollutant emissions, and promoting environmental legitimacy, thus creating an ecological city that is harmonious with the environment and social economy. The development of networked technology has brought many opportunities for production and human life. With the emergence of environmental problems, they have become a global problem, especially severe in Chinese cities. Therefore, in order to achieve rapid socio-economic development, it is necessary to introduce an intelligent environment that is conducive to building green cities. The research on data mining based on neural network-based smart environmental protection big data was of great help in improving the data mining performance of smart environmental protection platforms.

Keywords: Data Mining · Smart Environmental Protection · Neural Network · Big Data

1 Introduction

Considering the actual needs of environmental enterprises at present, the knowledge framework for the development of environmental systems should include the following contents: application systems, environmental information resource systems, infrastructure systems and frameworks, standard regulations systems, information security systems, and operation and maintenance management systems. The application system is the core of this framework. The application system is the core. The development and utilization of information resources are the foundation. Infrastructure is the support, standards, and norms.

A large number of scholars have conducted corresponding research on data mining. Sayah Z summarises the semantics in big data and summarises the various work done on energy savings in smart homes and cities. An efficient architecture is developed

J. C. Hung et al. (Eds.): FC 2023, LNEE 1131, pp. 59–65, 2024.
https://doi.org/10.1007/978-981-99-9299-7_8

based on ontology, big data and the synergy of multiple intelligences. The results of the experiments will provide a useful reference for the development of big data-based architectures. The research highlights that when used jointly, the two will provide a reliable energy-efficient system for smart cities. The results of the research improve energy efficiency and environmental protection while ensuring occupant comfort [1]. Suncare P's extensive research on the Internet of Things and cloud technology made it possible for people to accumulate a large amount of data generated from this heterogeneous environment and transform it into valuable knowledge through the use of data mining techniques. In addition, these generated knowledge would play a crucial role in intelligent decision-making, system performance improvement, and resource and service optimization management. An overview of cloud assisted Internet of Things big data mining systems was also introduced to better understand the importance of data mining in Internet of Things environments [2]. Suma V conducted a detailed analysis of the Indian e-commerce market using data mining technology to predict demand for upgraded e-commerce products. It also analyzed the impact of actual factors on demand and the impact of variables on demand. This analysis studied real datasets from three randomly selected e-commerce websites. Data was collected, processed, and validated using powerful algorithms. Based on the analysis results, it was found that despite the influence of various customer behavior and market factors, the proposed method could predict with high accuracy [3]. Although the above literature studies have proposed technical solutions for data mining problems, the application scope of the technology is relatively narrow.

With the continuous development of science and technology, the district and county environmental protection departments have established a basic computer software, hardware, and internet application environment. The platform construction level is low, and the problem of duplicate construction is very prominent; at the same time, there are also problems of data dispersion, information fragmentation, application fragmentation and service fragmentation among environmental management departments; at present, there is a common phenomenon of "seeing" and "using" "two skins" in China's environmental data, which makes it difficult to support the formulation of environmental management policies. This article would compare and analyze various algorithms of neural networks in big data data mining, so that readers have a broad understanding of data mining technology.

2 Data Mining, Smart Environmental Protection Big Data, and Neural Networks

2.1 Data Mining

Data mining technology:

Data mining technology refers to the step of discovering existing problems (knowledge base), which is an information processing process to automatically discover and extract specific relationships between data from large unstructured data sets. The main task of data mining is to find relationships between objects, cluster analysis, classification analysis, and anomaly analysis [4]. Rough neural network algorithm is a hybrid of rough set algorithm and artificial neural network algorithm. In data mining, when there is uncertainty or fuzziness in the data structure, rough neural network algorithm is often used.

2.2 Smart Environmental Protection

The concept of smart environment is a new method based on digitization and information technology to monitor the urban environment in real-time, while improving the automation, intelligence, and unmanned nature of the urban environment, improving the quality of pollution monitoring data, reducing the use of human and material resources, and fundamentally reducing the prevention of environmental overload caused by various reasons [5].

The challenges faced by implementing smart environmental protection in the construction of ecological cities:

Different Internet of Things products cannot work together well and there is no standardization. However, collecting information at the level of registering an intelligent environment is impossible without Internet of Things technology. Intelligence in a standardized and unified environment provides a unified data collection method and enables effective sharing of environmental data across the entire Internet of Things infrastructure [6, 7].

In order to create an intelligent green city, fast and accurate data processing technology is needed. Modern environmental construction cannot be separated from other building plans, so when establishing a data processing layer, it is necessary to fully consider data support from other fields in order to form a reasonable data processing mode [8]. In this mode, establishing a dynamic data management process is the only way to generate data with high reference value. When constructing a system oriented towards the environmental protection level of smart cities, the first step is to introduce the division of environmental protection areas, and clarify the resource distribution, ecological status, and environmental protection difficulties in this environmental protection task [9].

Overall architecture:

The intelligent environmental protection system would use a reasonable and scientific design concept to comprehensively evaluate the current situation of information technology, business needs, and data requirements. With the help of advanced technologies such as the Internet of Things, cloud computing, SOA (Service-Oriented Architecture), GIS (Geographic Information System), and data storage, it would create a comprehensive platform with intelligent and reasonable architecture, clear hierarchy, and aesthetic practicality. The system consists of four levels: monitoring system, data source, block service, and intelligent application, as shown in Fig. 1.

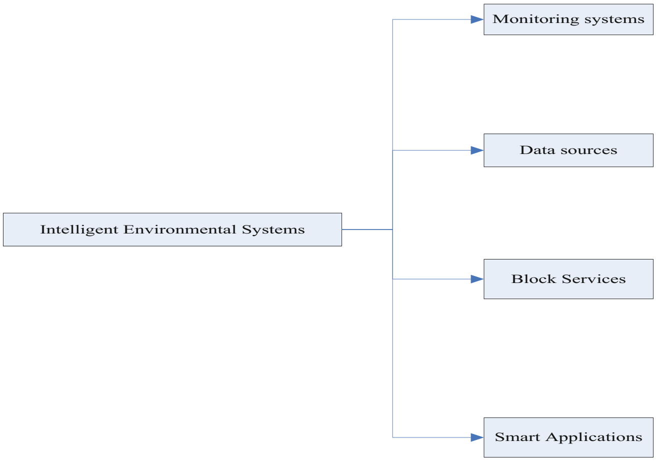

Fig. 1. Architecture of Intelligent Environmental Protection System

2.3 Neural Networks

Artificial Neural Networks (ANNs) have developed rapidly since their inception in the 1930s. This is because they can perform large-scale parallel processing on data and have high fault tolerance and transferability. The BP neural network algorithm is a multi-layer feedforward network trained based on the principle of error backpropagation. BP neural networks are one of the widely used artificial neural networks because they do not require prior knowledge of the mathematical formulas describing the dependency relationships of the problem.

Learning algorithms:

When there is a gap between the network output and the expected result, E represents the output error. The hidden layer is as follows:

$$\mathrm{E} = \frac{1}{2}\sum_{k=1}^{l}\left[d_k - f(net_k)\right]^2 = \frac{1}{2}\sum_{k=1}^{l}\left[d_k - f\left(\sum_{j=0}^{m} w_{jk}y_i\right)\right]^2 \quad (1)$$

The output layer is as follows:

$$E = \frac{1}{2}\sum_{k=1}^{l}\left\{d_k - f\left[\sum_{j=0}^{m} w_{jk}f\left(net_j\right)\right]\right\}^2 \quad (2)$$

According to Formula (1), the expansion of network error to the output layer is only related to d_k and W_{jk}. To reduce the error, it is necessary to modify the weights of neurons y_i [10].

It is assumed that the network has h hidden layers and m neurons. y_1, y_2,….., y_i represent the output of the hidden layer in the network. W_1, W_2,…, W_{h+1} are weights. The weight modification formula is as follows:

The output layer is as follows:

$$\Delta w_{jk}^{h+1} = \alpha \chi_k^{h+1} y_j^h = \alpha (d_k - O_k) O_k (1 - O_k) y_i^h \tag{3}$$

The hth hidden layer is as follows:

$$\Delta \mathrm{W}_{ij}^h = \alpha \chi_j^h y_i^{h-1} = \alpha \left(\sum_{k=1}^{l} \chi_k^o W_{jk}^{h+1} \right) y_j^h \left(1 - y_j^h \right) y_j^{h-1} \tag{4}$$

From the recursion, the formula can be obtained as follows:

$$\Delta W_{pq}^1 = \alpha \left(\sum_{r=1}^{m} \chi_r^2 W_{qr}^2 \right) y_q^1 \left(1 - y_q^1 \right) x_p \tag{5}$$

3 Experimental Evaluation of Data Mining Using Neural Networks

In order to effectively compare the advantages of different neural network classification algorithms in data mining, three types of corn seeds were selected from the database, and a subset of SVM (Support Vector Machine) neural network algorithm, BP (Back Propagation) neural network algorithm, and ELM (Extreme Learning Machine) neural network algorithm was used. In order to clean up the samples, 200 records were extracted from the database. Each record contained attributes such as grain density, length, and width, and the corn seed type was represented by binary digits {1, 2, 3}.

(1) SVM neural network algorithm

After using the loading function to import all the raw data into a dedicated storage file, in order to ensure the effective universality of the data, only 80% of the datasets were randomly selected, totaling 160 datasets. They were organized as training sets in the statistical model, while the remaining 20% of the datasets were organized as test sets. After setting the network parameters, the SVMpredict function was directly used to run the experiment. In order to make the experiment more realistic and reliable, a total of five iterations were conducted.

Table 1. Experimental results under SVM neural network algorithm

	First	Second	Third	Fourth	Fifth	Average
Accuracy rate(%)	89.848	94.858	97.2	89.811	94.918	93.327
Modelling time(s)	0.5703	0.122	0.1847	0.1696	0.1952	0.24836

From the analysis of experimental data results in Table 1, it could be shown that the average accuracy and model establishment time of the training set based on SVM neural network algorithm were 93.327% and 0.24836 s. Considering the high reliability of the research equipment used in practice, the data mining technology based on SVM algorithm had high credibility.

(2) BP neural network algorithm

The program for classifying the number of corn seeds using the BP neural network algorithm was basically the same as the program for classifying the number of corn seeds using the SVM neural network algorithm. After normalizing the training data, the number of output and input neurons was 3 and 7, respectively, which corresponded to a total of 3 corn seed types and 7 corn seed features for each type. When designing the BP neural network, the hidden layer node used an S-shaped tangent transfer function, and the output layer node selected a pure linear Newton transfer function. The BP algorithm was used as a learning function. After setting the network parameters separately, the experiment trained the BP neural network using training and simulation functions, and conducted simulation tests on it. The test results after five iterations are shown in Table 2.

Table 2. Experimental results under BP neural network algorithm

	First	Second	Third	Fourth	Fifth	Average
Accuracy rate (%)	82.281	77.831	82.494	87.685	85.546	83.167
Modelling time (s)	1.8007	1.7311	1.7668	1.7864	1.7412	1.7652

After careful statistical analysis of the experimental results, it could be concluded that the average accuracy and average simulation time of using the BP neural network algorithm to classify the dataset were 83.167% and 1.7652, respectively.

(3) ELM neural network algorithm

The method of saving raw data was to import all the raw data into a separate file called a seed, which was created using the load function. In order to ensure the universality of the data, only 80% of the dataset was randomly selected, which meant a total of 160 records. A statistical model was constructed from it as the training set, and the remaining records were used as the test set. The mapminmax function of MATLAB was used to normalize the training dataset, thus resulting in three output and seven input neurons, and representing the number of seed types and attribute features.

In order to create and train ELM neural network, elmtrain function was used to extract attributes and assign them to activation function. After assigning network parameters one by one, the ELM predict function was used to model and test the ELM neural network. In order to make testing more realistic and reliable, it was necessary to repeat testing as much as possible. A total of five iterations were conducted, and the final results are shown in Table 3.

Further statistical analysis of the experimental results shows that the average accuracy of the ELM algorithm was 96.192%, and the average modeling time was 0.119 s.

In this study, several neural network-based classification algorithms were analyzed. The experimental results showed that the ELM algorithm based on neural networks had strong modeling ability and high accuracy. The BP neural network algorithm had the lowest accuracy and requires more time to establish the model. The overall performance

Table 3. Experimental results under ELM neural network algorithm

	First	Second	Third	Fourth	Fifth	Average
Accuracy rate(%)	97.537	95.049	97.495	95.038	95.843	96.192
Modelling time(s)	0.0497	0.4177	0.0476	0.0428	0.0357	0.119

of SVM algorithm and ELM algorithm was vastly different. However, the ELM neural network algorithm used fewer nodes in the hidden layer, resulting in better performance.

4 Conclusions

In the transition from digital environmental protection to intelligent environmental protection, the most important thing is to start from the original digital environmental protection and focus on strengthening the perception layer and building an intelligent layer. Firstly, Internet of Things technology must build sensor systems that can perceive environmental parameters in real-time and adaptively. Secondly, it is necessary to apply intelligent information technologies such as cloud services and fuzzy perception, integrate existing information resources, build high-speed computer systems, storage capabilities, and parallel systems for processing environmental information, and ultimately build support platforms and information services for various applications of "smart environmental protection".

References

1. Sayah, Z., Kazar, O., Lejdel, B., et al.: An intelligent system for energy management in smart cities based on big data and ontology. Smart Sustain. Built Environ. **10**(2), 169–192 (2021)
2. Sunhare, P., Chowdhary, R.R., Chattopadhyay, M.K.: Internet of things and data mining: an application oriented survey. J. King Saud Univ.-Comput. Inf. Sci. **34**(6), 3569–3590 (2022)
3. Suma, V., Hills, S.M.: Data mining based prediction of demand in Indian market for refurbished electronics. J. Soft Comput. Paradigm (JSCP) **2**(02), 101–110 (2020)
4. Arcinas, M.M., Sajja, G.S., Asif, S., et al.: Role of data mining in education for improving students performance for social change. Turk. J. Physiother. Rehabil. **32**(3), 6519–6526 (2021)
5. Shneiderman, B.: Human-centered artificial intelligence: three fresh ideas. AIS Trans. Human-Comput. Interact. **12**(3), 109–124 (2020)
6. Ford, A.T., Ågerstrand, M., Brooks, B.W., et al.: The role of behavioral ecotoxicology in environmental protection. Environ. Sci. Technol. **55**(9), 5620–5628 (2021)
7. Gonzalez, R.A., Ferro, R.E., Liberona, D.: Government and governance in intelligent cities, smart transportation study case in Bogotá Colombia. Ain Shams Eng. J. **11**(1), 25–34 (2020)
8. Inderwildi, O., Zhang, C., Wang, X., et al.: The impact of intelligent cyber-physical systems on the decarbonization of energy. Energy Environ. Sci. **13**(3), 744–771 (2020)
9. Rymarczyk, J.: Technologies, opportunities and challenges of the industrial revolution 4.0: theoretical considerations. Entrepre. Bus. Econ. Rev. **8**(1), 185–198 (2020)
10. Samek, W., Montavon, G., Lapuschkin, S., et al.: Explaining deep neural networks and beyond: a review of methods and applications. Proc. IEEE **109**(3), 247–278 (2021)