Artificial Intelligence Technologies and Applications
C. Chen (Ed.)

doi:10.3233/FAIA231327

Research on Hydrogeology and Water Resources Evaluation Based on Big Data Technology

Gaofeng XU, Jie SUN[1]
Shanghai Urban Construction Vocational College, Shanghai, 201415, China

Abstract: With the development of science and technology, the amount of hydrological information data has undergone tremendous growth. How to fully utilize these large-scale data to support decision-making and water resource evaluation is currently a problem faced by scientists and society. In the traditional process of water resource allocation, the calculation of reservoir water balance results in a huge amount of abandoned water, resulting in the waste of a large amount of available water resources. By training Big data, the model self-learning can quickly analyze the characteristics of the river water level discharge relationship, fit the water level discharge relationship curve, and obtain the water balance, which reduces the number of flow measurements required to fit the water level discharge relationship curve by traditional methods, thus reducing errors, improving accuracy, and reducing time and economic costs.

Keywords: Water resource evaluation; Big data; water balance

1. Introduction

Water resource evaluation is a prerequisite and basis for comprehensive water conservancy planning [1-3], rational allocation of water resources, and scientific management, and has important significance for the safety of water resources and water ecology in river basins or regions. Due to the impact of global climate change and human activities on watershed or regional water resources, water resources face complex and diverse problems. Traditional water resource evaluation methods have a series of problems, such as distorted water resource quantity, separation of water quality and quantity evaluation, and separation of surface water and groundwater resource quantity evaluation. There is an urgent need for more scientific water resource evaluation methods to meet the needs of current strict water resource management systems. The water resources assessment based on the Hydrological model [4-6] can meet the needs of water resources management in the new situation and provides a new idea for water resources assessment in areas without data and changing environments. The traditional water resource evaluation methods focus on using mathematical statistical analysis methods. The subjective experience component in the evaluation process is large, and the work task is heavy, resulting in limited evaluation scope, mode separation, long cycles, and inaccurate handling of certain problems. The water resources assessment of the Hydrological model based on Big data technology [7-9] shows great advantages in terms

[1] Corresponding author: Jie SUN, Shanghai Urban Construction Vocational College, Shanghai, 201415, China; email: 13761710710@139.com

of systematic theoretical methods, practical efficiency of assessment process, and objectivity of assessment results, and is favored by hydrologists at home and abroad.

2. A model for Water resource Evaluation

Regional soil water resources are the average values of soil water resources from the perspective of water balance [10-12] in a watershed, reflecting the regional patterns of soil water. The differences in soil water resources are caused by different geomorphic units, lithology, and land use conditions in a watershed. The annual replenishment of soil water resources is used to represent the soil water resources under specific soil water conditions and land use conditions, and can also be used for regional water balance analysis. The annual replenishment of soil water resources under the condition of large groundwater table depth is calculated as follows:

$$SoW_{\mathrm{r}} = AnnPre + CondW - AnnPre_{\mathrm{int}} - Repre - Rechar \quad (1)$$

A univariate linear regression equation between the difference between precipitation and runoff and evaporation has the following relationship between the regression coefficient and the variance of the variables in the equation:

$$Regco = r_{xy}\frac{S_y}{S_x} \quad (2)$$

Land hydrological cycle simulation mainly controls the various substances present in sub-watersheds, which are basically input processes, while river convergence process simulation mainly controls the movement process of various substances present in river channels. The relevant model calculates the water cycle based on the water balance equation as follows:

$$SoW_{S,T} = SoW_{S,O} + \sum_{i=1}^{T} R_{day} - Q_{surf} - E_{cof} - W_{seep} - Q_{gw}) \quad (3)$$

In order to better process water resource data, we will preprocess the obtained data and obtain the relevant water resource dimension data. The preprocessing comparison chart is as follows:

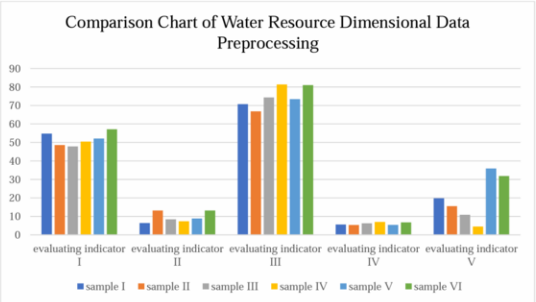

Figure 1. Comparison Chart of Water Resource Dimensional Data Preprocessing.

According to Equations 1-3 and Figure 1, we can get the total soil moisture content which is represented by the initial soil moisture content, time, total natural precipitation on the day, total surface water flow on the day, total water resource evaporation on the day, total soil moisture loss on the day, and total groundwater flow on the day. This algorithm model covers natural precipitation and various types of water flows, taking into account surface water flow, evaporation water flow, groundwater flow, and other aspects in the design, and integrating resources from various data channels.

3. Related Models Based on Big Data Technology

The tasks of data mining mainly include feature rule mining, identification rule mining, association (divided into Boolean and quantitative association) rule mining, classification rule mining, data clustering, prediction, trend rule mining, formula discovery, visualization, and data correction. Algorithms include inductive learning, genetic algorithms, neural networks, fuzzy mathematics, association analysis, rough sets, etc. Data mining enzyme results can be inputted in the form of tunnels, reports, logical formulas, and so on. The process of data mining includes data preprocessing, model establishment, mining implementation, rule extraction, knowledge interpretation, etc.

Data extraction is the process of extracting data from multiple databases or files and removing redundant data. For example, database software is used to jointly query statements for strongly labeled data. Data cleansing is to reprocess the extracted data, remove noise data and irrelevant data, and perform necessary interpolation and smoothing. For example, when reconstructing temporal and auditory sequences, it is necessary to obtain data with equal temporal and auditory intervals. Therefore, interpolation processing and removal of gross errors are necessary. Data conversion is the process of denoising monitoring data and converting it into the form required by data mining algorithms. The key to data preprocessing is to standardize the temporal and auditory Leopard data, reduce the size of the extracted dataset while maintaining the original data and integrity as much as possible, and improve the execution efficiency of the algorithm. In this paper, the data set is Discretization to remove irrelevant data. The calculation equation is as follows:

$$\chi^2 = \sum_{i=1}^{2} \sum_{j=1}^{k} \frac{Q_{ij} - W_{ij}}{W_{ij}} \tag{4}$$

The overall goal is to reduce the amount of abandoned water in the reservoir, improve water resource utilization efficiency, and maximize water supply. The key is to reduce the total amount of regional water shortage. The design objective function is as follows:

$$F_{\max} = F_1(x) + \alpha \mathrm{F}_2(x) + \beta \mathrm{F}_3(x) + \gamma \mathrm{F}_4(x) + \theta \mathrm{F}_5(x) + F_6(x) \tag{5}$$

4. Application of Big Data Technology to Hydrological Resources

The overall hydrological information processing architecture based on Big data sources is composed of the Internet of Things, water ecological monitoring stations, and Internet resources. Data collection of data sources is completed through remote upload, web crawler, and other methods.

The data storage module mainly stores the data processed by ETL technology in the HDFS distributed file system and HBase structured distributed storage system. This distributed file system improves the fault tolerance and availability of the data, provides support for file operation and storage, and can read and write in real-time and access randomly when necessary. The data is divided into blocks and Distributed storage on nodes in the HDFS cluster reduces the single-point performance limit of the database. It can not only provide high scalability but also achieve concurrent access. The distributed computing framework MapReduce helps us complete distributed computing programming on the Hadoop platform and complete the classification of water ecological data, and the Sqoop tool can complete the mutual transfer of data between Hadoop and relational databases. For example, by using Shell commands to backup structured data that exists in a relational database with structural specifications and does not require processing to a Hadoop system, data transfer is completed, thereby improving the speed of data acquisition.

The regional water resource sustainable carrying capacity prediction system and traditional prediction system studied in this article predict the sustainable carrying capacity of water resources in the same area, record the prediction time and accuracy of the prediction results, and analyze the performance of the two systems based on the obtained results. The comparative experimental results are as follows:

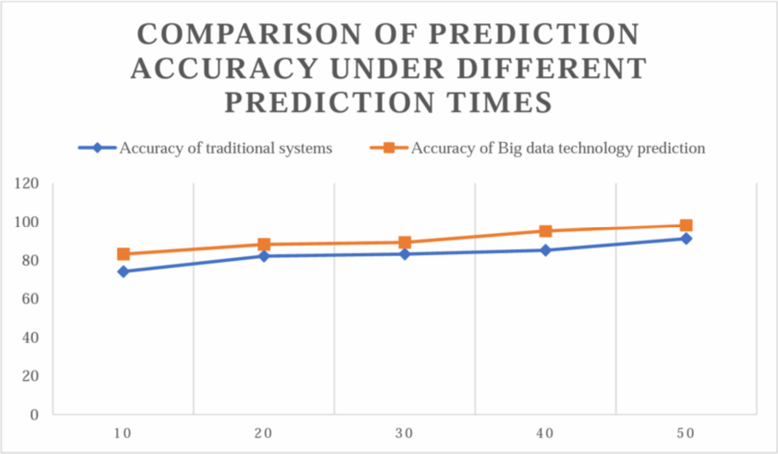

Figure 2. Comparison of prediction accuracy under different prediction times.

From the experimental results in Figure 2, we can see that using big data technology has improved accuracy compared to traditional methods. At the same time, as the sample data increases, the accuracy of each method also improves.

5. Conclusion

The research results show that the combination of Big data technology and water resources assessment can reduce noise, remove redundancy, and improve prediction accuracy and efficiency by preprocessing data before forecasting the original data series. Although there may be some errors in the difference between the predicted and measured hydrographs, algorithmic optimization should be carried out to minimize noise and improve prediction accuracy as much as possible. Through the improvement of the traditional model, a dynamic allocation model of water resources is proposed. Combining Big data technology, the model is solved by training Big data and cross validation. In the calculation, the uncertain incoming and outgoing water data are replaced by the storage capacity of the reservoir period, which avoids the uncertainty of the model and makes the model more reasonable and reliable. Finally, the test and comparative analysis are carried out.

Acknowledgments

This paper has received funding from Shanghai Urban Construction Vocational College (No. cjky202215). At the same time, this paper also received support from the Advanced Visiting Program for Key Teachers in the Yangtze River Delta.

References

[1] Liran Funaro, Orna Agmon Ben-Yehuda, and Assaf Schuster. 2019. Stochastic resource allocation. In Proceedings of the 15th ACM SIGPLAN/SIGOPS International Conference on Virtual Execution Environments (VEE 2019). Association for Computing Machinery, New York, NY, USA, 122–136.

[2] Mithun Chakraborty, Ayumi Igarashi, Warut Suksompong, and Yair Zick. 2021. Weighted Envy-freeness in Indivisible Item Allocation. ACM Trans. Econ. Comput. 9, 3, Article 18 (September 2021), 39 pages.

[3] Kushagra Kinger, Ajeet Singh, and Sanjaya Kumar Panda. 2022. Priority-Aware Resource Allocation Algorithm for Cloud Computing. In Proceedings of the 2022 Fourteenth International Conference on Contemporary Computing (IC3-2022). Association for Computing Machinery, New York, NY, USA, 168–174.

[4] Ogundepo Odunayo, Naveela N. Sookoo, Gautam Bathla, Anthony Cavallin, Bhaleka D. Persaud, Kathy Szigeti, Philippe Van Cappellen, and Jimmy Lin. 2021. Rescuing historical climate observations to support hydrological research: a case study of solar radiation data. In Proceedings of the 21st ACM Symposium on Document Engineering (DocEng '21). Association for Computing Machinery, New York, NY, USA, Article 19, 1–4.

[5] Lars Arge, Aaron Lowe, Svend C. Svendsen, and Pankaj K. Agarwal. 2023. 1D and 2D Flow Routing on a Terrain. ACM Trans. Spatial Algorithms Syst. 9, 1, Article 3 (March 2023), 39 pages.

[6] Amita Singh, Magnus Wiktorsson, Jannicke Baalsrud Hauge, and Seyoum Eshetu Birkie. 2022. A simulation-based participatory modelling framework for stakeholder involvement in urban logistics. In Proceedings of the Winter Simulation Conference (WSC '21). IEEE Press, Article 118, 1–12.

[7] Teng Haikun, Wang Shiying, and Xiao-Guang Yue. 2019. Application of Big Data Technology in Emergency Decision System. In Proceedings of the 2019 3rd International Workshop on Education, Big Data and Information Technology (EBDIT 2019). Association for Computing Machinery, New York, NY, USA, 93–97.

[8] Andrey Aleksandrovich Rodionov, Rabim Alikulovich Fayziev, and Saidakhror Saidakhmedovich Gulyamov. 2023. EXPERIENCE IN USING BIG DATA TECHNOLOGY FOR DIGITALIZATION OF INFORMATION. In Proceedings of the 6th International Conference on Future Networks & Distributed Systems (ICFNDS '22). Association for Computing Machinery, New York, NY, USA, 412–415.

[9] Alfredo Cuzzocrea and Paolo Ciancarini. 2021. SeDaSOMA: A Framework for Supporting Serendipitous, Data-As-A-Service-Oriented, Open Big Data Management and Analytics. In Proceedings of the 2021 5th International Conference on Cloud and Big Data Computing (ICCBDC '21). Association for Computing Machinery, New York, NY, USA, 1–7.

[10] Gerardo Aguilera and Kevin Blom. 2022. Eternals: Tackling the Large Scale Water Whirlpools. In ACM SIGGRAPH 2022 Talks (SIGGRAPH '22). Association for Computing Machinery, New York, NY, USA, Article 27, 1–2.

[11] Gerardo Aguilera and Kevin Blom. 2022. Eternals: Tackling the Large Scale Water Whirlpools. In ACM

SIGGRAPH 2022 Talks (SIGGRAPH '22). Association for Computing Machinery, New York, NY, USA, Article 27, 1–2.

[12] Yaseen A. Al-Mulla and Taif B. Al-Badi. 2022. IoT and RS Techniques for Enhancing Water Use Efficiency and Achieving Water Security. In Proceedings of the 2021 4th International Conference on Machine Learning and Machine Intelligence (MLMI '21). Association for Computing Machinery, New York, NY, USA, 83–88.