Research Article

Marine information database design

Lintao Li and Shanhong Zhu

School of Computer and Information Engineering, Xinxiang University, Henan, China
_____

ABSTRACT

In order to solve the problem of marine product data of multi-source heterogeneous and make users to manage a large number of complex ocean product data effectively, we design and implement a set of marine product information management system. First ,we introduces the system architecture, and then give exposition to each module in detail, and carry on the detailed analysis of the Marine data, aiming at the problems in marine product data such as the data inconsistency ,the gap and redundant data, we use time series of anomaly detection method based on cumulative variation, it can detect the abnormal point in the ocean data effectively, and modify abnormal points again, then guarantees The quality of the Marine data, thereby a set of perfect Marine product information management system is established, it makes the related department of Marine business process more scientific and standardized.

Key words: abnormal detection; data management; ocean data; information management system
_____

INTRODUCTION

As the country's sea growing demand and the rapid development of marine economy, the demand of oceanic administration management and service ability is presented in order to satisfy the demand of marine users in ocean data management and data analysis requirements, designing marine information management system is imperative, however, the marine data quality affects scientific decisions in information management system, currently there is no systematic marine data quality control methods ,generally uses the manual correction processing control to deal with a large number of marine data, this paper uses time series Detection technology based on cumulative abnormal variation[1]. Abnormal detection is carried out to collect the ocean data, the qualified data and the revised legal data are stored in the database to manage and display marine information.

The system architecture design
Marine information management system mainly includes data acquisition, abnormal detection, data preprocessing, data storage and management, the architecture of the system is as shown in figure 1.

SYSTEM FUNCTION MODULE DESIGN AND IMPLEMENTATIONDATA ACQUISITION MODULE
Data acquisition module includes meteorological sensor, hydrological sensors and biosensors weather sensors to collect meteorological data, including wind speed and direction, precipitation temperature and fog data; Hydrological sensor collects data including water temperature, wave current, water salinity, Sediment concentration and the suspended sediment, etc; Biological sensors collect phytoplankton and zooplankton benthic life Objects[2].

Through the analysis, original data exist with the following problems:
(1)Marine data exists empty value in certain fields.so we need these data transformation and integration, need intelligent fill to null data field
(2) Data of site information on each station remain the same in the structure, but the integrity and consistency of data is very poor

_____

(3) Similar data from different data tables have different data types that date is sometimes with the date type, sometimes with character type
(4) Marine data in each station more or less contain noise data, must be washed before into data warehouse

To sum up, the ocean information management system of the original data exists different consistency, data gaps, data redundancy issues, marine data doesn't directly used in subsequent data mining, marine data preprocessing is the precondition of data mining[3].
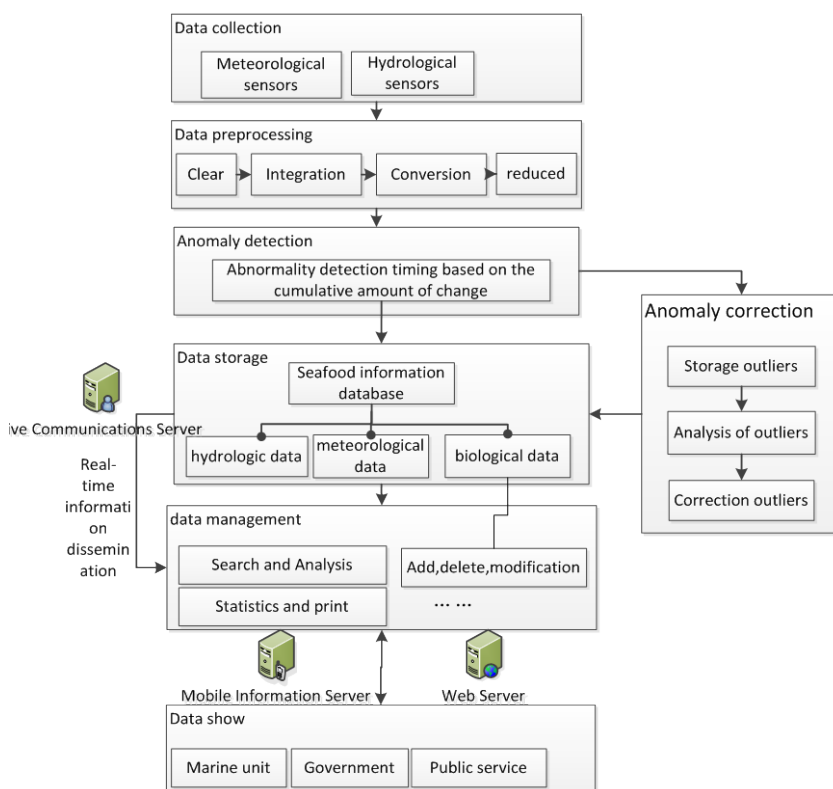

Fig.1 The architecture of information management system

DATA PREPROCESSING MODULE
Marine data mining is the application of data mining technology in the field of ocean, from a lot of incomplete, fuzzy, noisy random ocean data, extracting The implied, prior unknown and potentially useful ocean information and knowledge, due to marine ocean mainly represent as ocean time series, hence, ocean data mining is the mining of temporal data, according to statistics[4], in the whole process of data mining ,data preprocessing spend sixty percent or so, and then accounts of the digging take ten percent only, after data preprocessing, it not only can save a lot time and space, and the mining results can play a better decision-making and forecasting Part, as a result, the data preprocessing before entering the data warehouse is guaranteed[5].

Marine information management system of data preprocessing is mainly achieved through the four aspects of the work: data clean up, integration, transformation and reduction and so on:

(1)Data cleaning is the standardization before the data into the data warehouse, is a number check according to the integrity and consistency, data cleaning routines through filling in the missing value, light sliding noise data, recognition or remove outliers and resolving inconsistency, cleaning up the number" in order to achieve the following goals: main format standardization, abnormal data elimination, error correction, duplicate data.

(2)Data integration is to combine and unify data in multiple data sources, these data sources may include multiple database, data cube or general file.in data integration, the need to consider the problems include: data conflict, data loss and derivative data [6].

(3) Data transformation is to convert data into a form which is more suitable for data mining, the method that mainly includes the data type transformation of the data segment, data collection, data generalization, normalization transfer

_____

the data into the form to data mining.

(4) Data reduction is used to get data set reduction, as it is much less, but still is close to maintaining the integrity of the original data, the results with the results before reduction is the same or nearly the same.

ANOMALY DETECTION MODULE
Anomaly detection module mainly use time series based on total variation, anomaly detection methods to anomaly detect the marine data, to store for qualified data, to store on the abnormal point for the unqualified data and abnormal analysis, to select the appropriate correction method to modify abnormal points[7].

DATA MEMORY MODULE
Data storage module is mainly store legal data after time series of anomaly detection to ocean information database.

DATA MANAGEMENT MODULE
Data management module is through adding, removing, requiring, analyzing and statisticsing operations mainly to the marine meteorological, marine hydrology and marine life, etc[8].data importing function can increase vastly, can import data into database, can make it conveniently and efficiently to improve work efficiently, for query function, through the precise query and fuzzy query modes, Marine data query for all conditions, and the query result displays with chart.as shown in fig.2,3.

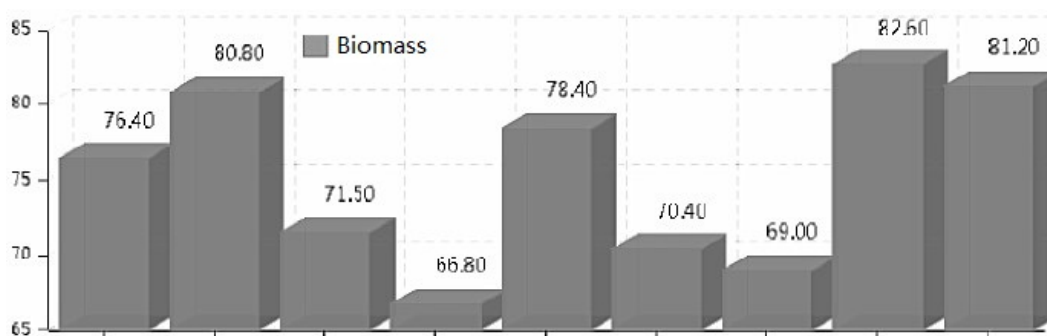| 1 | 3305 | -7.00 | 1980 | 1 | 5 |
2	3305	10.50	1980	2	5
3	3305	12.90	1980	3	5
4	3305	15.00	1980	4	5
5	3305	17.50	1980	5	5
6	3305	20.00	1980	6	5
7	3305	23.00	1980	9	5
8	3305	30.00	1980	8	5
9	3305	19.40	1980	7	5

Fig.2 Meteorological data management module

Fig.3 Biological data management module

Data display module
Data display module display the result of data analysis module by a variety of formats of chart. The client displaying include marine unit, public user, government users, displaying module and data management module are GIS server and dynamic information server to achieve real-time communication and display.

Data analysis functions are compared by comparing stations and for many years, to a sea Of a certain historical time trend and a normal Marine elements analysis, The rule of the Marine data summed up, and to give the user, for policy makers For help[9].

_____

ANOMALY DETECTION TECHNOLOGY

At present, the method for anomaly detection in time series, the method based on frequency, the method based on feature space and machine learning methods, but these methods were not applicable to Marine time series, due to the abnormal form is different, the time sequence of the abnormity can be divided into abnormal sequence, abnormal point and abnormal patterns, this article mainly aims at the characteristics of the ocean of time series data, using abnormal detection method based on total variation to detect abnormal point in the ocean of time series, this method involves an average amount of statistics, the fluctuation magnitude of time sequence of observations, this variety is the average of the difference in value adjacent observations:

$$MeanChange = \frac{|v_2 - v_1| + |v_3 - v_4| + \ldots + |v_n - v_{n-1}|}{n} \tag{1}$$

Among them, V_i Refers to the observed value, n refers to the number of observations

The conception of average variation on the basis of the cumulative variation is proposed and the concept of marine time series abnormal points have given the definition

Definition 1 :Marine time series abnormal point definition

Given a Marine time sequence $X=<x_1=(v_1,t_1),x2=(v_2,t_2),\ldots,xn=(v_n,t_n)>$,点 $xi(v_i,t_i)$represents the time sequence at this tithe moment of observation value is vi$<N_1,N_2,\ldots N_K>$ represents point xi of k neighbors sets, the observed value set is for$<v_1,v_2,\ldots,v_k>$, given threshold T, If the point x_i With the accumulation of two neighbor points variation is larger than T, then decision point xifor This period of an exception point in the ocean of time series, the decision condition is as follows:

$$\frac{w_1 \cdot |v_i - v_{N_1}| + w_2 \cdot |v_i - v_{N_2}| + \ldots + w_k \cdot |v_n - v_{n-1}|}{w_1 + w_2 + \ldots + w_k} > T \tag{2}$$

$<w_1,w_2,\ldots w_k>$ Refers to the weight vector in formula(2),Give each variable quantity different value, among which, the $<w_1,w_2,\ldots w_k>$ is named as $<1,2,\ldots,j>$,Usually, in the timeline, the closer to the neighbor points X_i ,the greater value assigned to it; Threshold T refers to a constant. The accumulation of Variation of T and the size of threshold T determine whether X_i Is an abnormal points.

This article detects the sea abnormal point using cumulative variation time series of anomaly detection method. As shown in figure 1,this method is to read the original data in the first place, and calculates the average data Volume, then traverses each data point, checks Point next to its neighbors, calculates the cumulative amount change Values, calculate threshold C according to the average variation.

Analysis and anomaly correction technology

The reasons of producing abnormal data are man-made factors in computing and wrong operating errors, or the data inherent property, the marine time sequence of point xi is judged as abnormal point according to definition 1, The Xi Point and the value of its neighbor points accumulation of variation is bigger, there are many causes, the possible reasons can be described as the following three categories combined with the characteristics of the ocean of time series,(1)due to errors in the process of inputting data (2)Natural factors (3)Other man-made factors，abnormal correction method mainly includes the following four points：manual correction on the basis of other data sources，estimation by the field Experts.

CONCLUSION

In this paper, a Marine information management system is developed based on Marine data Detailed analysis, we found that the marine data in information database has characteristic of the data inconsistent, data gaps and data redundancy. in order to better conclude and preprocess to the marine data, abnormal variation Point test method is proposed based on the cumulative of the time series of abnormal detection method and then analyze the abnormal points, and select the appropriate correction method Abnormal points for correction, provide concise and accurate Data for marine information database，ensure the quality of the Marine data, so as to improve the Marine information management system performance of the system. The system can quickly and efficiently manage the

_____

Marine data, fast, accurate and comprehensive professional management requirements, can provide scientific support to marine related department management.

REFERENCES

[1] Reorganization of Consumer Health Protection and Food Safety in Germany. FAO/WHO Global Forum of Food Safety Regulators . **2004**.

[2] Jensen,H.H.,LUnnevehr.HACCP in Pork Processing: Cost and Benefits. *Economic of HACCP Studies of Cost and Benefits* .**1999**.

[3] Thompson M,SylviaG,Morrissey M T.Seafoodtracing ability in the united states: current trends system design and potential application. *Comprehensive Reviews in Food Service Food Safety* .**2005**.

[4] United States Government Accountability Office. USDA and FDA Need to Better Ensure Prompt and Complete Recalls of Potentially Unsafe Food. http://www.gao.gov/ cgi-bin/getrpt?GAO-05-5 . **2004**.

[5GruerP, Hilaire V, et al. Ames formal framework for multivalent systems analysis and design. *Expert Systems With Applications* .**2002**.

[6] Gaurav Banga,Peter Druschel.Measuring the Capacity of a Web Server.*Proceedings of the USENIX Symposium on Internet Technologies and Systems* .**2001**.

[7] New Features in IIS 4.0. MSDN Library Visual Studio 6.0, 1998 Mercury Interactive. Automatic Testing of Internet and InternetApplication. *Database and Network Journal* .**1996**.

[8] Shouhong Wang,Analyzing Agents for Electronic Commerce. *Information Systems* .**1999**.

[9] Ann Balough,Building a Knowledge Center for Records and Information Management. *Records & Information Management Report* .**2001**.