Research on the optimization of data mining error system based on dynamic information

Qingru Sui[1, a]

[1]Changchun University of Science and Technology, Changchun City of Jilin Province, 130600, China

[a]sqr1979@21cn.com

Keywords: Dynamic information; Data mining; System optimization

Abstract. At present most of the data mining technology can only on the basis of the past to produce large amounts of data analysis processing, it is difficult to meet the fast pace of society people expect of knowledge novelty. Therefore, this article puts forward the dynamic information of the data mining technology, and on this basis on the analysis of the dynamic information collection, dynamic information, the concrete analysis of the data window. In order to further optimize the error system based on dynamic information of data mining, with sliding window as the dynamic information collection and analysis platform, and on the basis of data mining, and add the evaluation based on data mining test link in management, in order to help to the development of error data mining system.

The development of the error data mining system

With the rapid development of computer network technology, now entered the era of big data, big data rapid development and popularization, and humans began to truly realized data the vastness of the world. In the face of such a huge data resources, a large number of numerical analysis become the bottleneck of data development, there is an urgent need for new data analysis methods and techniques, to be able to use the new achievement of the development of the information technology, these huge data resources translate into actual production and life has a huge help information and knowledge, to provide support for scientific decision for us. For preliminary to overcome and solve this problem, a new data analysis method and technology arises at the historic moment, it is the data mining system.

It is because of the data mining technology in the field of big data analysis and processing advantages, both in academia and the information technology sector began to gradually use the related technology to expose data of the underlying information. Based on this, all walks of life to build a large amount of related knowledge base and database,according to these get tens of thousands of increase the amount of data, hope unearthed a large number of useful data in the database, can is to guide and promote role for future work. Data mining technology, while in theory and application aspects have made remarkable achievements, but there are still many inadequacy and defect of the theory and application technology, is worth further study.[1] We live in a every day there are a lot of new information, analysis of the information society, therefore, we need through the data mining technology to deal with a lot of new data, to meet the need of actual production and life. But, we have ignored an important question: at present most of the data mining technology can only on the basis of the past to produce large amounts of data analysis processing, to extract which contain in-depth information, but when such information was discovered, it may have been out of date, it certainly can't satisfy the fast pace of people in the society of knowledge of novelty expectations. Therefore, based on the past, present and future data under the dynamic information of data mining technology is undoubtedly one of the ways to break through this bottleneck.

The data mining process under the dynamic information

Dynamic Information Collection. Dynamic information acquisition is to point to obtain dynamic data collection.[1] Different from the static data sets, dynamic data sets include not only

information has occurred in the past, also contain the happening now and future is about to happen. Dynamic information collection is not done overnight, it is a step by step, have layers. In the first place in the past there have been information collection is in the past and current data set by extracting once completed; Secondly for is happening now and future of impending information collection is in the past, current and future data set by ceaselessly extract complete it step by step. Due to the dynamic information collection and data acquisition in a different way, so in order to prevent data of discontinuity, we use the sliding window as dynamic information collection platform. as displayed in Fig. 1[2]. At the same time, to show the characteristics of the dynamic data is different from the static data, this paper will take the time variable as the sliding window units.
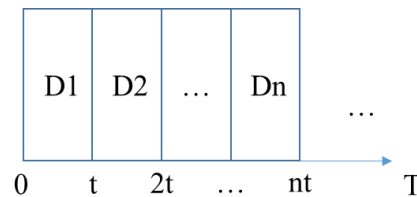

Fig. 1 The Sliding Window

In the data mining system based on dynamic information under, will generate time T to t($t \cong n$, $n \cong 1$)for the time segment, data di for data segments in each period is divided into Dn. each data window corresponds to a Dn, Where n is the maximum value of the data window. At some point T, have a certain data set D = {D1, D2,... Dn} to enter the data window maximum value for t data window, every once in a while t0, the data window will be moved forward from m ($1 \leqslant m \leqslant n$), among them m refers to the distance m data window, also referred to as the sliding window. With sliding window as the specific process of dynamic information collection platform visible figure 2, the m = 2.
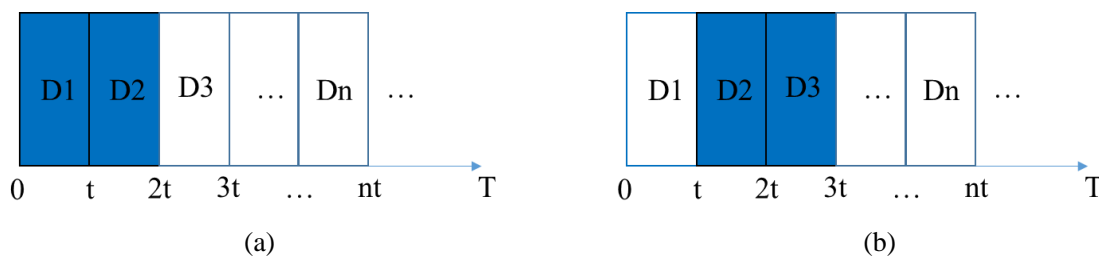

(a)   (b)

Fig.2 With Sliding Window as the Dynamic Process of Information Collection Platform

The Analysis of the Dynamic Information. Generally deal with those in the time scale of database is stored on the correlation between the weaker data sets or there is no correlation between discrete data sets [3]. Therefore, in order to store processing on time scale correlation strong dynamic data set, Must establish a new type of database, based on the data mining system under the dynamic information of the database, so that, through its rapid retrieval function find people need to analyze the various types of data. In this premise, based on the dynamic information of data mining technology can not only for static data analysis and processing, can also be for a variety of dynamic real-time data analysis and processing, to help people get the latest information. Among them, the analysis of the dynamic information is mainly refers to the analysis of the dynamic data sets collected processing, mainly includes the following aspects: eliminating noise, make up for the loss of data, transform the data type, extract the data characteristics, to reduce the data dimension, etc. [4]. In dynamic analysis of the data set during the processing of data can be used by code method, but pay attention to the update of real-time data, moreover, can also be overlapped window technology combined with the regulations of the data, data in sliding window to parse the dynamic information, in order to avoid omissions boundary data.

Dynamic Data Window. Dynamic data window is a platform for dealing with limited dynamic data, the sliding window is one of these types. Dynamic data window is the most commonly used method is the application of the method of cluster analysis to the overlapping Windows mobile

technology [5]. For example, a finite dynamic data collection Z = {z1, z2,... , zn} RC divided into (2 k - 1) a data window,when the limited dynamic data collection overlap appears in the adjacent two data window, for calculation of the two overlapping data again, so on, until the end of the whole data set. Compared with the simple clustering analysis method, this method not only preserves the advantages of clustering analysis, and made up for its shortcomings, can quickly find and calculate the type smaller clustering,so, we don't have to worry about the omission of adjacent data window boundary and density point contribution inconsistent problem [6].

The error data mining system based on dynamic information optimization research

The error data mining system based on dynamic information. In the process of data mining based on dynamic information, dynamic information collection and analysis is the key to prevent the formation of discrete data, the data set and is happening now in the past has smooth transition between data sets, as well as the data sets that are happening now and the future is going to happen smooth transition between the data sets. However, based on the data mining also has flexibility under the dynamic information, this is because the dynamic information are characteristic of volatility, universality, timeliness, etc. Under the assumption based on dynamic information of the data in the data mining system for d_i, the current point in time T, may occur in the data mining based on dynamic information under the following process:

(1) If the number is negative integers, based on the data mining system under the dynamic information in T - n (n > 0) generated by the data set is formed by the past there have been data sets, for D1.

(2) If the number of positive integers, based on the data mining system under the dynamic information generated in T time data set is formed by data sets that are happening now, remember to D2.

(3) If number is not the integer, based on the data mining system under the dynamic information in time T + n (n > 0) generated by the data set is formed by the data set is going to happen in the future, remember to D3.

(4) Based on dynamic information of data mining is the data mining system based on dynamic information of D1, D2, D3, all of the data analysis process, to dig out the latest information and has the value of knowledge.

As a result, the data mining based on dynamic information is gathered in the past there have been data set, data sets that are happening now and data sets is going to happen in the future, in order to analyze the data behind the latest information and valuable knowledge. This has to do with data mining based on static information, compared to extract may be outdated information, has a huge advantage. Next, therefore, we need to explore in depth research based on the data mining process under the dynamic information structure, As shown in figure 3. Can be seen from the diagram, based on the dynamic information of the process of data mining structure can be roughly divided into dynamic information, dynamic data processing, set the target data and evaluation data management process. Based on the dynamic information of the flow chart of data mining, evaluation of management data is one of the important link, because through the evaluation of management data mining result of the previous round, can detect the deficiency in the process of data mining, the next round of data mining can be modified.
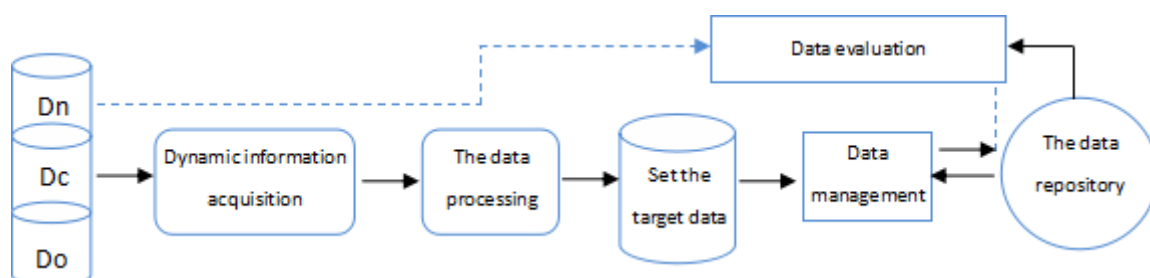

Fig. 3 Based on the Data Mining Process Under the Dynamic Information Structure

Need to conquer the problem of data mining based on dynamic information is dynamic data

acquisition and processing, which prevent the formation of discrete data. In order to solve this problem, This article adopted the cluster analysis method, will be overlapping Windows mobile technology and its application, to prevent the occurrence of discrete data and real-time dynamic data acquisition, analysis, and processing. Because of data mining based on dynamic information in the process of operation, the data mining system based on dynamic information under at work, too, that the data set is going to happen in the future is constantly expanding, therefore, data analysis and data automatically update processing efficiency improving technology progress is to ensure a data mining based on dynamic information under the foundation of success.

Wrong data mining system optimization. Through the analysis of data mining based on dynamic information of knowable, based on the dynamic information of data mining technology has yet to be promoted, the error data mining system based on dynamic information under also remains to be further optimized. In view of this, in this paper, on the basis of data mining, add mining test this link, in order to the error data mining system based on dynamic information of the optimization of the help.

Data Mining. In data mining based on dynamic information, people for analysis and processing of different data types have different requirements. In the past has happened information collection, because of the lag has occurred, so people on the efficiency of mining processing requirement is not high, the general data mining techniques can be used to set a maximum mining time interval value of the δ, when the time interval to the δ, the last end of the data mining and start the next data mining; But for now is happening and will happen in the future information collection, due to new information degree is higher, so people also higher requirement for the efficiency of the mining process, must adopt a new data mining technology to set an interval mining value σ, The interval σ must be within the information of the actual results. To sum up, based on the dynamic information of data mining to the processing of different data types can have different methods, but the efficiency of mining processing must be in a higher level.

Operation process of data mining based on dynamic information, collection of data that are happening now and the future is going to happen the data set is constantly expanding, the real-time data is constantly growing. Every moment in dynamic data window will appear s new data window, so in order to make every data d_i in each period of time D_n can not be missing data, in the data window s as the basic unit of the dynamic data window, to analyze the specific process of data mining, as shown in figure 4: Data d_i

① The first window for 0 ~ s;
② The second window for v ~ s+ v(0 <v<s);
③ The third window for s ~ 2s;
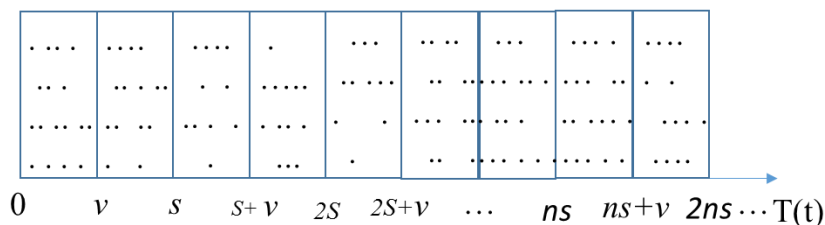④ The fourth window for s + v ~ 2s+ v;
etc..

Fig.4　Dynamic data window schemes based on overlapping Windows mobile technology

Overlap in the adjacent two data window, for calculation of the two overlapping data again. Over time, the data window to move forward and have overlapping part of the data window, repeat calculation and so on. In figure 4, v is the specific time period, the value of v should not be too long, also should not be too short, because time is too long, the data is more; The time too short, overlap of data window is more, increasing the time of data processing. For example of dynamic data in the data window d_i data set D = {D1, D2,... Dn}, the time interval of each data segment is t, as data window s + 1 as the basic unit of the dynamic data window,and v=s; From each data set D adjacent

data extracted feature for processing.

The Mining Test

Mining test is under the dynamic information of data mining in the test data mining result of the previous round, it can detect the deficiency in the process of data mining, in order to improve the next round of data mining. If everything is normal, you can carry out data mining; if test results appear some problems, you need to inspection on the part of data mining, in order to ensure the normal operation of subsequent data mining. In the process of real production practice, test data mining is very difficult, because of data mining based on dynamic information contains data type is more, analysis and processing methods are not unified. But that does not represent data mining test does not need to, on the contrary, we must combine test data mining and data mining, in order to for data mining based on dynamic information of error system is optimized.

Conclusion

At present most of the data mining technology can only on the basis of the past to produce large amounts of data analysis processing, it is difficult to meet the fast pace of society people expect of knowledge novelty. So, in this paper, based on the past, present and future data under the dynamic information of data mining technology. In based on the analysis of the data mining process under the dynamic information, the acquisition of dynamic information, the analysis of the dynamic information, dynamic data window and error data mining system optimization has carried on the detailed analysis: In order to prevent data of discontinuity, let the sliding window as dynamic information collection platform; The analysis of the dynamic information including noise cancellation, make up for the loss of data, transform data types, extract the data characteristics, to reduce the data dimension, etc.; Dynamic data window becomes a bridge dynamic information collection and the analysis of the dynamic information; The evaluation of data management is based on the data mining process under the dynamic information structure is an important part of the,so on the basis of data mining, add mining tests can help error data mining system based on dynamic information of optimization.

Reference

[1] Zhou Jinge. Data mining based on the theory of the fault elimination error system optimization method and application research [D]. Guangdong, guangdong university of technology, 2014:27.

[2] PengRan. Based on dynamic data mining of power plant thermal system operation optimization method research [D]. Beijing: north China electric power university, 2012:19.

[3] ChengLin Niu,   Jizhen Liu, Jianqiang Li.Yang data mining algorithm based on fuzzy partition in the application of power plant combustion optimization system [J]. Computer and applied chemistry, 2008 (7) : 902-906.

[4] Zhou xiaoyun, QinXiong pie. Database lock table based on lightweight data mining method optimization [J]. Computer engineering and application, 2012 (36) : 16-20.

[5] Hai-feng Huang, Ke-hong Zhang , Zhang hong. Power system dynamic information database key technology [J]. Journal of computer applications, 2011 (6) : 1681-1684.

[6] MingXin Teng ,Zhongyang Xiong, Yu-fen Zhang. Dynamic data mining research [J]. Journal of computer applications, 2008 (6) : 160-162.