

Online Monitoring of Indexing Technology Based on Parallel Computing for Large Data Stream

Wei Hu¹, Jialin Zhang²

¹School of Physical Education, Jiangxi Institute of Fashion Technology, Nanchang, 330201, China

²School of Physical Education, Jiujiang University, Jiujiang, 332005, China

wei_hu2014@yeah.net

Keywords: Online monitoring, Layers index, Discrete attributes, Associated data; Large data.

Abstract. In order to realize the large data online monitoring facing the rural sports project, this paper proposes a hierarchical index structure and its on-line rule matching algorithm, the attribute set on large data stream is divided into discrete attributes by the layers of index. The use of online matching method and the multi-level index of the associated data relation table carry out filter and optimize effective for monitoring results. Through the on-line monitoring data analysis, its experimental result shows that compared with the classical on-line monitoring method, it goes through the hierarchical index structure without reducing the premise of accuracy, it significantly enhances the monitoring efficiency of large data flow, which provides a theoretical reference for the study of large data online monitoring method.

Introduction

In the rapid development of computer network and large data aspect era, data shows the characteristics of more and more attributes heterogeneous, and the so-called heterogeneous large data stream refers to a variety of different types attributes large data flow, which collectively refer to as the heterogeneous large data stream [1,2]. In the application of large-scale data online monitoring analysis, people will adopt registered large-scale monitoring rules in heterogeneous large data stream [3]. In order to satisfy all the rules, we must use the minimum computational overhead; at the same time, in order to solve the huge large data monitoring rules, the key is to improve the performance of rule detection.

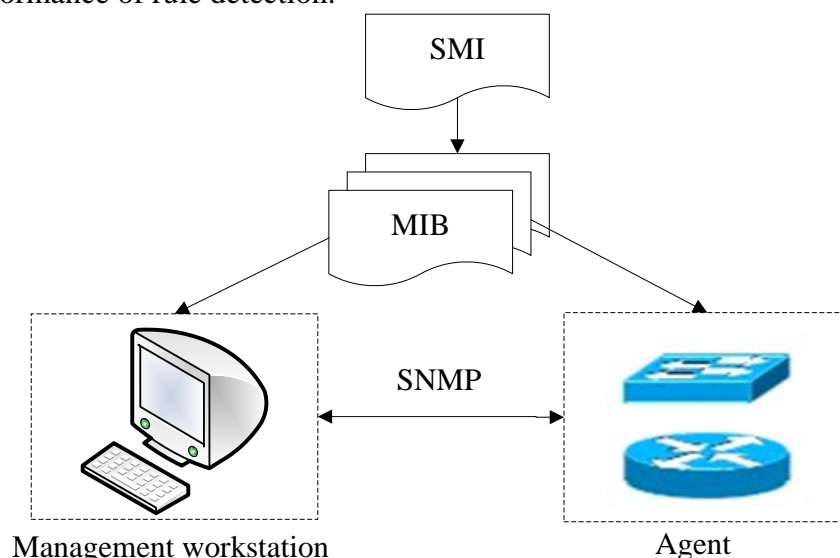


Fig.1: The frame structure of large data on-line monitoring

Typical layer index structure is SNMP protocols, which is respectively structure management information (SMI), management information base (MIB) and SNMP [4,5]. SMI and MIB define communication using data types between SNMP agent and workstation, and SNMP defines that between management station and agent carry out communication protocol and protocol data unit format of data interaction.

Design of Rural Sports Project Large Data Stream Hierarchical Index Online Monitoring Algorithm

Because national rural sports belong to large data stream monitoring in the monitoring of construction projects, we carry out content monitoring for the heterogeneous data stream, firstly we need to register a large number of monitoring rules in data stream; secondly we should pay attention to match the key, we will real-time reach each tuples and registration monitoring rules in data stream to carry on matching [6-8]. For the definition of monitoring rules and data stream tuples, we use hierarchical index to carry on data on-line monitoring as shown in Figure 2.

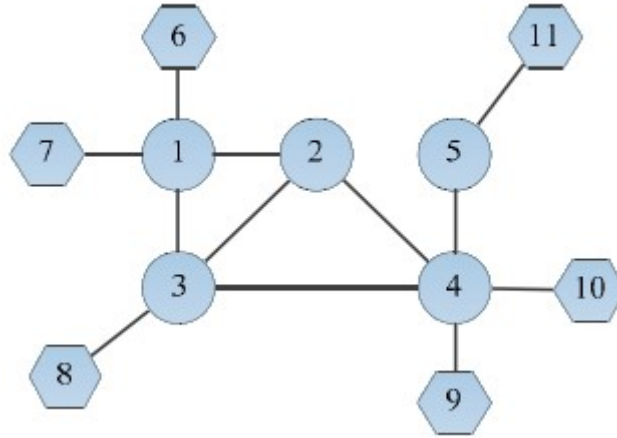


Fig.2: Hierarchical index diagram

The main storage data structure of hierarchical index is adjacency matrix and adjacency list, and the network topology is a typical sparse matrix, making the queue and stack to store the topological adjacency list structure, assumption hierarchical index structure is shown in Figure 3.

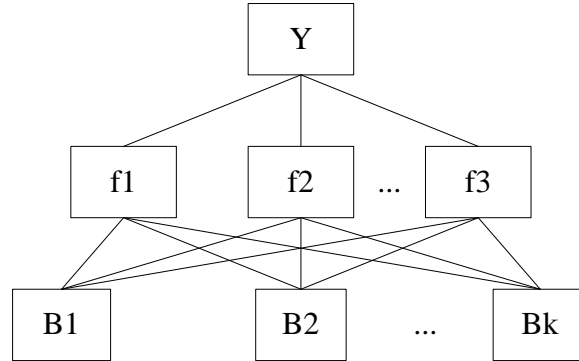


Fig.3: Hierarchical index structure diagram

As shown in Figure 3, f shows the hierarchical index function. According to the underlying rules, online monitoring index target Y carries out sort that is B_1, B_2, \dots, B_k , in which j -th project test index of k layer is

$$B_{1i}, B_{2i}, \dots, B_{ji} \quad (i = 1, 2, \dots, j) \quad (1)$$

The hierarchical index of N layer is

$$\begin{aligned} N_1 &: f_1 B_{11} + f_2 B_{12} + \dots + f_k B_{1j} \\ N_2 &: f_1 B_{21} + f_2 B_{22} + \dots + f_k B_{2j} \\ &\dots \\ N_k &: f_1 B_{k1} + f_2 B_{k2} + \dots + f_k B_{kj} \end{aligned} \quad (2)$$

Among them, the total target weight of x -th index in N layer is

$$\sum_{y=1}^l f_x n_{yx} \quad (3)$$

Then each level index can be written as

$$\begin{aligned} \sum_{x=1}^l f_x n_{1x} &= n_1 \\ \sum_{x=1}^l f_x n_{2x} &= n_2 \\ &\vdots \\ \sum_{x=1}^l f_x n_{kx} &= n_k \end{aligned} \quad (4)$$

Hypothesis level single ranking index is KL_i and random consistency index is GL_i , then the same ratio of the level total order is

$$KH = \frac{f_1 KL_1 + f_2 KL_2 + \cdots + f_k KL_k}{f_1 GL_1 + f_2 GL_2 + \cdots + f_m GL_m} \quad (5)$$

When $KH < 0.1$, we can think that large data online monitoring index is optimal; if it is not optimal, we will readjust the consistency ratio [9]. The algorithm is as follows:

FOR all rural sports project x DO

 a ← rule First: TreapSearch(x).

 IF a is not empty THEN

 H

 second: Search(x; res);

 ELSE

 x ← Get Next Tuple x;

 END IF

 IF res is not empty THEN

 res ← Combine(res; r).

 ELSE

 x ← Get Next Tuple x;

 END IF

END FOR;

.....

Study on the Large Data Rural Sports Item Index Hierarchical Analysis Simulation Experimental

In order to verify the effectiveness and reliability of the large data hierarchical index online monitoring algorithm, this paper uses artificial data set to carry on experiment for the monitoring rules component separately, effective distribution governs the relationship between the predicate and monitoring rules [10,11]. Based on the keyword cable of content, real-time multimedia data captured from gateway flow is testing data stream, and then extracting attributes and dimensions have been selected, the schematic diagram of its protocol operation is shown in Figure 3.

By monitoring algorithm takes the heterogeneous data stream form to extract the contents, its experiment can be deployed in e Duo2.66GHz CPU processor, 2GB memory, AS5 Linux machines [12]. The experiment is the key to observe the time overhead of monitoring rules calculation, in which taking an arbitrary data stream tuples complete the computation cost time of all the monitoring rules that is defined as time overhead.

It shows the data online monitoring node interface. Online database scanning end sends data for the database, it goes through the hierarchical structure to carry on n node distribution, and then according to different index to determine the search rules, it will realize the data multi-level index monitoring [13]. The real-time data monitoring results are shown in Figure 4.

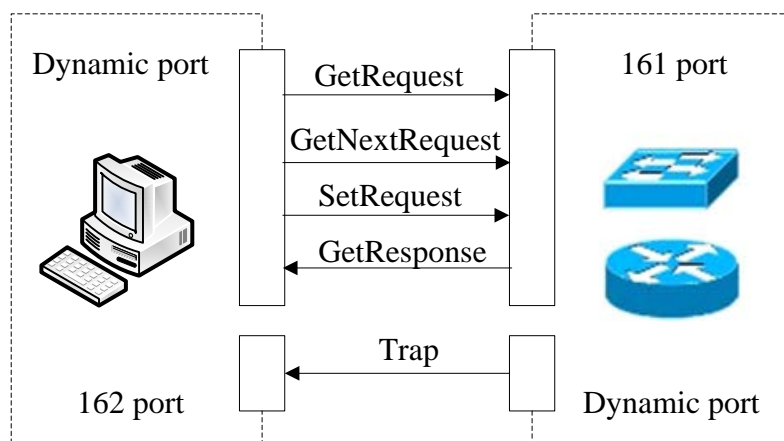


Fig.3: The schematic diagram of SNMP protocol operation

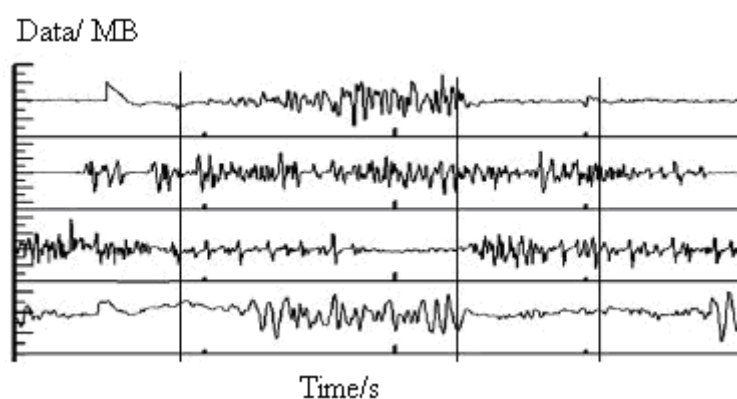


Fig.4: Multi-level index flow monitoring

Figure 5 shows the real-time data curve of multi-level index traffic monitoring. It can be seen from the chart that in accordance with the multi-level index online monitoring rules, rural sports carry out real-time monitoring in four levels, including the number of project construction, sports project type, the number of sports exercise and sports project equipment, in which general data flow monitoring curve is shown in Figure 5.

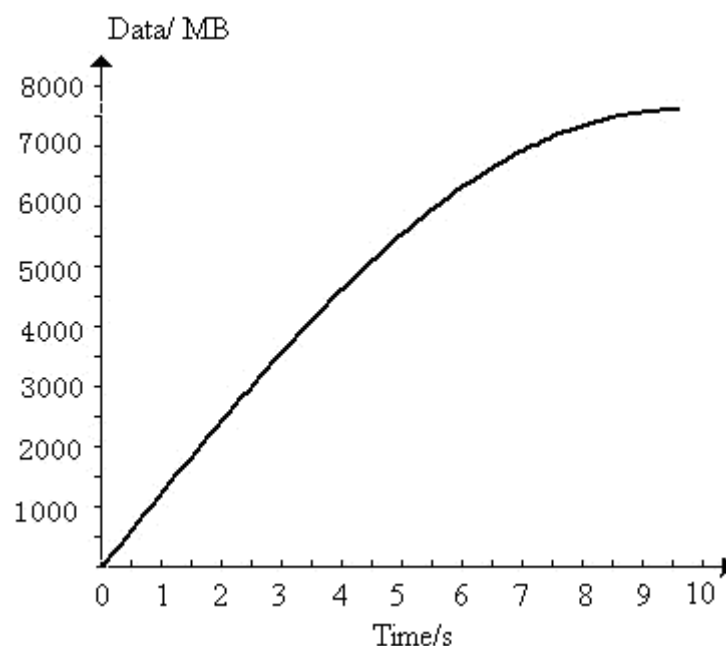


Fig.5: Large data flow monitoring real-time curve

Figure 5 shows the overall real-time curve of large data flow monitoring. It can be seen from the chart that along with the increase of time, the monitoring amounts of large data flow is growing, there is no big fluctuations, so as to verify the validity and reliability of the algorithm.

Table 1: Flow test record list

Test time	Time consumption	The number of sports
2014-06-08	8:03:06	829
2014-06-08	9:12:06	930
2014-06-08	8:35:06	825
2014-06-08	8:22:06	841
2014-06-08	8:11:06	838
2014-06-08	8:48:06	828

Table 1 shows the large data monitoring recording list in the movement project, data record includes rural sports project construction, sports project type, the number of movement project, sports project equipment and funds and so on data. For each data monitoring, they have adopted hierarchical index recording method, wherein consumption time and the number of sports project are shown in table 1. As can be seen from table 1, the sports consumption time of different project is different, but the same movement projects close to the overall time consumption. Under the level index method, consumption time is integrity lower, which meets the demand of large data on-line monitoring.

Summary

- (1) According to large data characteristic of rural construction movement training project, this paper proposes the data online monitoring method of hierarchical index, it has realized the index monitoring of large data different attribute level, which will improve the correlation degree of on-line monitoring data.
- (2) Using the online matching and multi-level indexing algorithm, this paper designs the sort and index problem of different data flow on-line monitoring realizing the online real-time monitoring of large data, and can obtain the real-time curve of large data monitoring through artificial data flow monitoring.
- (3) By the calculation results, it can be seen that the on-line monitoring method of multi-level index can guarantee the accuracy of data on-line monitoring, and the monitoring efficiency is also high, which is a new large data online monitoring method.

References

- [1] Q.L. Chen, Z.Y. Zhang, F. Wang. Study of access control model orienting multimedia social network. Journal of Xi'an Electronic and Science University, 2014 (6): 65-69.
- [2] D.H. Wang, A. Wang. Study of the efficiency of power analysis attacks According to the block cipher S box. Journal of Shandong University, 2014 (2): 36-39.
- [3] L. Wang. Parallel computing techniques. Information technology, 2012(10): 95-96.
- [4] R.M. Chen. The strategy of challenges, value and coping of large data age. Mobile communication, 2012(17): 63-64.
- [5] X.Q.Gong, C.Q.Jin, X.L.Wang. Data intensive scientific and engineering: needs and challenges. Chinese Journal of computers, 2012(8): 91-92.
- [6] X.J. Yang. Parallel computing sixty years. Computer engineering and science, 2012(8): 62-63.
- [7] Y. Zhou, X.W. Lu, C.T. Cheng. Analysis of high dimensional data stream canonical correlation parallel computing method in irregular flow. Journal of software, 2012(5): 42-43.
- [8] Y.Q. Kai, Z.F. Zhao, J. Fang, Q. Ma. The mass data real-time processing method according to high speed data flow. Chinese Journal of computers, 2012(3): 56-57.
- [9] J.Y. Liang, J.W. Gao, Y. Chang. The semi supervised learning research progress. Journal of Shanxi University, 2011(4): 31-32.

- [10] X.L. Tang, M. Han. The semi supervised learning method based on extreme learning machine. Journal of Dalian University of Technology, 2010(5): 21-23.
- [11] H. Li. Application of semi supervised learning and its data mining. Computer knowledge and technology, 2010(27): 101-104.
- [12] R. Liu, H.Y. Li. The semi supervised learning research and application. Software guide, 2010(8): 52-53.
- [13] W.J. Chen. Review of semi supervised learning research. Computer knowledge and technology, 2011(16): 75-76.