A Method of Collecting Data With Times Tamp in Internet

Shexuebing

Jiangxi Institute of Technology, department of digital technology,Nanchang,China

shexuebing1982@sohu.com

Keywords: Internet of Things;Timewindow;Timestamp; Data collection

Abstract. With the development of the Internet .this article introduces a method of time window,that can be used to collect as much legit data as possible via creating new time windows，inserting data into the time windows and destroying time windows. The number of time windows maintained depends on practical applications. The new coming data will be inserted into existing time windows according to its timestamp and source mark. If fail to insert，we then destroy some of the time windows and creatie new time windows to hold the new comers. Before destruction，all the data in the time window should be inserted into other time windows . The process of destroying and creating makes sure the time window is kept moving forward.

Introduction

Applications of computater occupy an important position in the Internet and cloud, inevitably, with a large number of them dealing with data, data collection, data analysis, and data processing. Therefore, to improve production efficiency,and to obtain considerable economic benefits, dealing with data accurately and efficiently has become a key factor[1]. Therefore ,there is a new focus on data collection algorithm. This algorithm puts forward some prospects : constantly transmitting data to the data source, how to collect as much information as possible, ensuring that the data collected from different sources of data, and keeping the latest data.

Take the sensing data in intelligent building network as an. If one sensor collects a data point at each time and sends the data to the data server, considering that many sensors, each sensor can collect data in does not stop, so as to give the data time stamped and sensor tags. We consider the following problem in computational aspects behind: in a formula in the calculation of the data from different sensor. To solve these problems, can collect links in the data, excluding pseudo data, the legitimate data for each algorithm to sort out an array or a queue, so as to achieve the purpose of calculating link.

The problem

The following is the one room in the intelligent building with the calculation of electric energy consumption data to illustrate the problem. The installation of a plurality of sensors in a room, lighting electricity data were collected, air conditioning electricity data and electric power data, then the total electric data adding them to get the room. From the sensor data is a structure, at least including electric data, sensor data acquisition time and acquisition of the data of the number of sensor. At the same time, in order to satisfy the rationality and the network transmission delay time calculation, two two in this paper agree above three were obtained from different sensor structure data of sensor data acquisition time difference is no more than 5 seconds to perform the addition. If there is a time difference of two data for more than 5 seconds, you need according to the actual situation of the requirements of the decision is to give up time on the first data or not to take any action to wait for the new data come together[2].

Then the following a situation may exist in the real scene to discuss issues: lighting electricity at time t came to L1 data, in t+t1 (T < T1) time pass over the data L2. Then we must make a choice, is to retain the data L1, or retention data or L2, while retaining the L1, L2? If the

retention of L1 obviously, defects, If the electricity data AC1 and electric power data A1 in t+t2 air conditioning (t+5 < T2 < t+t1 +5) moment came, obviously the data cannot be added, Because three data does not meet the time of the agreement, L1 data than AC1 and A1 was more than 5 seconds. If the retention of L2, this problem will be smoothly done or easily solved, total power consumption of T = L2+AC1+A1. But if hastily abandon L1, it may cause the waste of data, Because the data L1 is not used, then (t – 5, t + 5) could not get results this time period. Because, if in T3 (T < T3 < t+5) moment, electric power data AC2 and electrical data A2 for air conditioning. In this case, the total electricity consumption should have 2 results, respectively is T1=L1+AC2+A2 and T2=L2+AC1+A1. Of course, the data requirements are not stringent conditions, the retention of L2 from L1 is no problem, but in the data demanding circumstances, can not be arbitrarily discarded L1, But at the same time, the retention of L1 and L2 by using a method, used for back calculation. This paper is also retained L1 and L2 provide a solution, namely time window method[3].

Dynamic data structure

We take time window as an object-oriented language class. First, there is a list or array used to hold a data structure. Then have a interval field. This field is used to limit the effect of the length of the time window, which is any the time window of the two structures, the difference of currentTime values do not exceed the value of interval.

The value of the interval is based on the actual situation can be adjusted. If in fact the data accuracy is relatively high, is involved in computing the data time difference is small, i.e. as far as possible to ensure that data involved is almost at the same moment, You can put the interval value is set too small, can be a millisecond. Conversely, if the amount of data is relatively small, such as one minute intervals will generate a number of data, then the value can be set to more than. In addition, attention is drawn to the interval value settings will have another effect that influence the time window, the moving speed. Because once the new value can not meet the time window data time requirements, it will destroy the old time window and create a time window, the new window of time that contain data in time is relatively new, this is the time window moving speed. In conclusion, the interval value is smaller, the faster moving time window.

In addition, according to the actual need, can be a count field to add in each time window, this field is used to structure data has a number of markers in a time window. If the number of each time window containing the data of fixed, then count can with each structure of data coming from increased, when the count value reaches a preset number means that the data registration is completed, can put the time window of data submitted to the processing unit for processing. If each time window containing the number of the data is not fixed, the count field is not required, but according to the need to add their own.

In this paper, each structure contains two fields: currentTime and ID. In other scenarios can be added according to the needs of other fields. CurrentTime is used to mark the time attribute data of the structure, the structure of data is the time of the birth of. This field can not only mark the time attribute data of the structure to which it belongs, in time of need can also be based on the currentTime field to sort all the data in a time window. This paper uses this method[4].

As for ID, is used to represent the structure of data is generated by which sensor, this field can guarantee the existence of structure data each time window are from different sensors, no more data with a sensor of the situation there is a window of time, can lead to incorrect results.

There is a need to explain, structure of data in a time window can be disorderly, can also be ordered, can use currentTime to sort, can also be sorted by ID. These are to be adjusted according to the actual situation. To a time ordered time window into the data structure is very convenient, can

be used directly ordered insertion method for data structures. In this paper, the following agreement:the currentTime sequence of the sort of structure, to identify the unique structure of data according to the structure of ID. Another field of ID function is very obvious: mark the source of this structure of data. This ensures that all data structure in each time window from different data sources.

Time window operation

As mentioned above, the time window is a holding structure of data class, create a time window is to instantiate the class.

Step1 Now the false with 2 time window TW1 and TW2, the initial state of the two time windows are empty, no element. Now the first structure elements S1 arrival. Because of the previous two time windows are empty, then obviously the S1 will be successfully inserted into the time window in TW1. The initial time of S1 arrival will initialize the TW1, then the structure elements after the arrival of the Si must be greater than S1 currentTime in time, and less than currentTime+interval[5].

Step2 Second structure elements S2 arrival, find the right position to insert. Following insertion algorithm:

(1) First try to insert the S2 into the TW1. To determine the structure elements have the same sensorId Sj according to the S2 sensorId (the case S1 is S1) exists in the TW1. The judge will produce two results.

(2) S2 does not exist in TW1. Then the next step is to determine whether the S2 cur-rentTime satisfies the above conditions, obviously, this time on legitimacy will produce two results.

(3) S2 can be inserted into the TW1, note that S2 was inserted into the TW1 should pay attention to according to the time order of the sequence is inserted into the appropriate location.

(4) Unfortunately, S2 currentTime does not meet the time requirement, then S2 cannot be inserted into the TW1. However, because the TW2 is empty, then S2 could be successfully inserted into TW 2.

(5) the S2 exists in the TW1, then the TW2 is empty, the S1 could be successfully inserted into TW 2. Obviously, the head element TW2 is S2.

Discussed below two time windows are not empty, this possibility is relatively complex, involving the time window of destruction and reconstruction. There are elements of S11, assuming TW1 S12,???, S1m; there are elements of S21, S22,??? S2n. Suppose the structure elements of Sk. According to the Convention, the first attempt to Sk to TW1 to insert, if not successful try to insert TW 2. Insertion algorithm are as follows:

(6) whether there are sensorId and Sk a S1i in the judgment of TW1 sensorId's.

(7) TW1 does not come from the same sensor data, that all sensorId sensor-Id and TW1 Sk which is not the same. So, then, judge Sk currentTime meets S11.currentTime currentTime < S11.currentTime+interval < Sk

(8) the Sk time requirements are satisfied, then, put the Sk in accordance with its time is inserted into the appropriate location TW1.

(9)if Sk time does not meet the requirements of TW1, it took Sk to test TW2, to see whether it can smoothly into the TW2. Process and test the TW1 test is the same, First determine whether Sk sensorId only in TW 2, followed by judge Sk currentTime is suitable for the TW2. If the above two conditions of Sk are satisfied, then the Sk will go into TW2. If the Sk does not meet any one of these conditions, then carry out the following steps: the destruction of the window that TW1 and TW 2 smaller time (i.e., comparison of S11.currentTime and S21. current-Time), the head element

in an actual application has to allow the head element TW1 in the early hours of TW2 time, and then create a new time window TW3, finally, insert the Sk into the TW3.

(10) the existence of a a S1i TW1 in the sensorId field, sensorId field and the Sk S1i of the same, then go to step (9).

We now introduce the destruction of an old time window and a new process. Because according to the original design, the time of TW1 (TW1 S11 currentTime) as early as the TW2 time, then destroyed is TW1, then a new TW3,. From the first element of S11 TW1 began to carry out the following steps. Check whether the S11 sensor-Id already exists in TW 2, and S11. currentTime meets S21. currentTime < S11. < S21. currentTime+interval currentTime, If the above two conditions are met, then insert the S11 into the correct position of TW2; if you do not meet these two conditions in any one, then delete the S11, to give up the structure elements. For all structure elements in the TW1, in accordance with the above steps, so as to judge whether the S1i inserted into the TW 2, to meet the conditions of the S11, according to the currentTime field they insert them into the correct position for the TW2; For those not fully condition is S1i, direct delete them.

To complete the processing of all S1i, TW1 has become an empty time window, then it becomes TW3. So that it can accept the structure of new elements. Needs to be pointed out is, in the process of concrete is a bit different. If Sk cannot be successfully inserted into TW1 and TW 2, the TW1 theory into TW3, and then insert the Sk into the first element of TW1 position. The destruction of TW1 on the TW1 in addition to the first element (Sk) of all the elements are processed, whether they can be inserted into the TW 2.

But this method will produce another problem: in the former TW1 in order to add to the end position of the TW2. After completion of this step can take TW1 cleared. Next, the mobile TW 2 original structure to TW1. This is a mobile process includes two steps: first the original elements according to their TW 2 TW 2 in order to add to the TW1, add after the completion of these elements in turn can be removed from the TW 2.

Conclusions

This paper introducd the time window concept and related operations, including the new time windows, data insertion time window, as well as the destruction time window. By the time window algorithm to deal with the time stamp data, can be calculated link back to provide as much as possible in time also legitimate newer data.

References

[1] tangyujia. Design of energy management system based on Internet of things [J]. Computer applications and software,2011,28,（12）:161.

[2] zhangjunhu. IVDQP—The Internet of things interest vector based distributed query processing method [J]. Computer applications and software,2011,28（11）：207.

[3] yanweiming. Data structure [M]. Beijing: Tsinghua University press.1997.

[4] Donald Ervin Knuth. The Art of Computer Programming Vol2：Funda-mental Algorithms[M]. New Jersey：Addison-Wesley,2012.

[5] Donald Ervin Knuth. The Art of Computer Programming Vol3：Sorting and Searching[M]. New Jersey：Addison-Wesley，2011