© 2015 IJSRSET | Volume 1 | Issue 6 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

A Survey on Prediction of Missing Sensor Data Using Association Rule

Hitarth Chauhan*, Bakul Panchal

Department of Information Technology, L. D. Engineering College, Ahmedabad, Gujarat, India

ABSTRACT

Missing values is major problem in sensor network. Currently we have many existing approach to predict missing values in stream of data. But for pre fetched existing data we can't use such techniques. So while querying in such data will lead to wrong results. So in this paper we will try to predict such missing data in existing sensor data using association rule mining techniques.

Keywords: Window Association Rule Mining, K-nearest Neighbour Estimation, WSN, Data Reduction Mechanism, Data Mining, Sensor Data

I. INTRODUCTION

Currently there are many applications working on sensors. Sensors are now not just limited to weather forecasting. It is now used in many mobile devices and also many health care devices also uses sensors. At every second very large amount of sensor data are gaining generated. But gathering data from senor have many hurdle. As most of the time sensors are working to track peripheral environment it also faces many weather disturbances. It may also face power failure. Because of such reasons sensor data will always have some missing values. And when we try to query such missing data then gathered results will not be accurate. So we need some mechanism to retrieve those missing data.

We can always request such missing sensor data again but it will work only on data continues stream of data. For data with are already gathered this action will not work. So in this paper we will review some techniques to predict these missing data of sensor network stream and also to predicting such data from previously gathered sensor data.

II. METHODS AND MATERIAL

LITERATURE REVIEW

A. Determining Missing Values in Dimension Incomplete Databases using Spatial-Temporal Correlation Techniques [1]

In this paper author has provided technique to predict missing data of pre fetched data using association rule. This technique uses techniques like WARM (Window Association Rule Mining) and AKE (Applying K-nearest Neighbour Estimation). Querying dimension incomplete databases could lead to obtaining incomplete results. Considering this limitation this paper proposes to incorporate the above avoidance methods as a part of searching dimension incomplete databases and also proposes newer version to the existing WARM method. The advantage of the proposed approach is that the result of the user query will always have complete and accurate data.

This paper proposes to the above avoidance methods as a part of searching dimension incomplete databases. The advantage of the proposed approach is that the result of the user query will always have complete and accurate data.

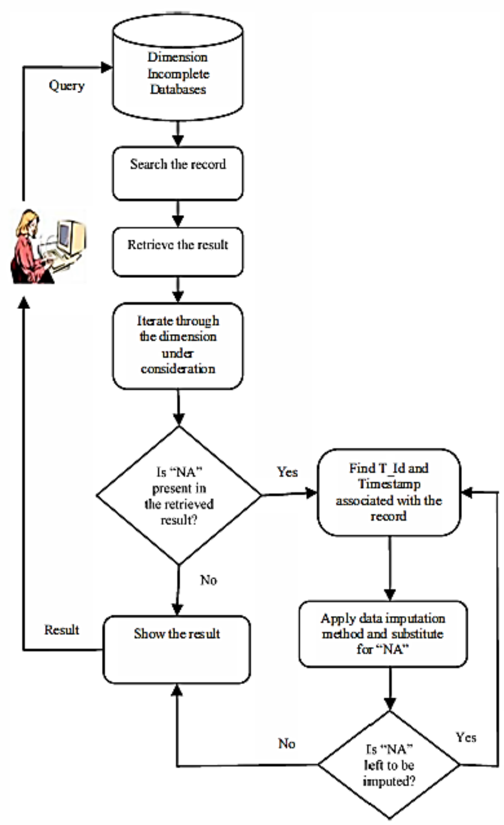

Figure 1: System Flow

Following are the steps to be carried out for Dimension Incomplete Databases:-

Step1 : Query the dimension incomplete databases
Step2: Search the records as per the user query.
Step3: Retrieve the result.
Step4: Iterate through the dimension under consideration within the retrieved result to check if any of its value is having "NA".
Step5: If "NA" is not present in the retrieved value then it indicates that the dataset or the data required for a given query is complete so we can directly go to step7.
Step6: If "NA" is present in the retrieved result then it indicates that our retrieved result set contains missing values. In that case we will have to follow the below steps.

a. Find the T_Id and Timestamp associated with the record having attribute values as "NA". *T_Id has been included for simplicity as the timestamp format can vary from user to user.
b. Apply any of the data imputation method suitable for the wireless sensor data. In our case we will be using WARM, two newer variations (Max-WARM and Pattern-WARM) of existing WARM method which takes spatio-temporal relations into consideration as our data imputation method. After imputation, substitute the estimated value in place of "NA".
c. If any more "NA" is left to be imputed then repeat step 6, else go to step7.

Step7: Provide the requester with the required data as the output for the query.
Other also used K – nearest neighbor techniques for the same purpose. With all techniques prediction capacity was 100%.

B. Using Data Mining to Estimate Missing Sensor Data [2]

In this paper author has provided technique to predict missing data from stream of data coming from sensors. In this research we propose a power-aware technique, called WARM (Window Association Rule Mining). An estimation of the missing value(s) is performed by using the values available at the sensors relating to the MS through association rule mining.

In this paper they are using relations exists between two sensor nodes. They will try to create association rule between each pair of sensors from previous set of data. And more the pair has same results more their association will get stronger. And for this process they have used data structure like Buffer and Cube. Every time they will not process all previously gathered data they will create one buffer structure and store previously gathered association detail in the buffer for each pair. And after every round of data extraction the buffer value will be updated.

In their proposed algorithm they use checkBuffer(), update(), and estimateValue(). checkBuffer() will check for missing value. If it finds missing value it will try to estimate it by applying estimateValue(). And after estimating value they will update value of buffer again using update() method.

C. Using Data Mining to Estimate Missing Sensor Data [3]

In this paper also author has provided technique to predict missing data from stream of data coming from sensors. They have tried to account for relationships among sensors and simultaneously, incorporate the time factor making the estimation process computationally aware of the relative relevance of each data round in the data stream. This paper was improvement upon previous one only. They have tried to include the freshness in predicting missing sensor data. This paper uses and follows three below mentioned processes.

a. Incorporate the temporal aspect into association rules and estimation
b. Compact data streams and allow a large history to appropriately influence sensor rules
c. Guarantee retrievability of original data from its compact form

To include freshness in the estimation process they have introduce one multiplier factor in update() function. They will multiply the freshness factor to the estimated association. And on each round they will increment the freshness factor by 1. So every time we will try to estimate value the resent value will always have higher contribution in estimation process. But this approach will also have limitation of limitation of system memory. After certain round the value of multiplication factor will go very high such that system will not be able to store such huge value.

D. An Enhanced Data Reduction Mechanism to Gather Data from Missing Sensor Association Rules [4]

In this paper, we have tried to enhance the algorithm proposed by Azzedine et. al by removing more redundancy between sensor activities. We have focused only on sensor association rules that they have mentioned, which is as below.

Let S = {s1, s2,. . ., sn} be the set of sensors in a sensor network. Let time is divided into equal sized slots {t1, t2, . . . tm} such that ti+1 – ti = _, where _ is size of each time slot. This = tm – t1 is the historical period of the data defined during the data extraction process is the historical period of the data defined during the data extraction process. A set of sensors P = {s1, s2,. . . sk} is called a pattern of sensors. An epoch is a couple E =(Ets, P) such that all sensors in the pattern P detect events within the same time slots (Ets), and a sensor database DS(also called behavioral data) is defined to be a set of such epochs(Table I). An epoch E(Ets, P) is said to support a pattern P1, if P1 _ P.

TABLE I: DATABASE OF EPOCHS. *TS*: TIME, S: SENSORS

Ts	Epoch
1	s_1, s_2, s_5
2	s_1, s_3, s_5
3	s_2, s_4

WSN can be assumed to be consist of a set of sensors nodes S = {s1, s2, s3,. . .,sn}. Time is divided into slots of equal size _. Each sensor has a buffer B ,one entry for each time slot. At the end of every time slot, each sensor sets the corresponding buffer entry to denote whether an event has been detected at that time slot. For example, a sensor buffer with seven time slots may look like {0, 0, 1, 1, 0, 0, 1 }, which shows that events have been detected by the sensor during 3rd, 4th and 7th time slots.

III. CONCLUSION

As, all others have suggested that best way to get sensor missing values is to predict it from existing values rather than requesting sensor to pass same value again. We have looked process of estimating missing sensor data for already gathered data. In that process we can include freshness factor to make prediction process more accurate. For that we need to change some process of including freshness in association rule. This paper will help people to understand reason for requirement of data prediction technique in sensor network data.

IV. REFERENCES

[1] Sneha Arjun Dhargalkar, A.D. Bapat "Determining Missing Values in Dimension Incomplete Databases using Spatial-Temporal Correlation Techniques", In 2014, IEEE

[2] Mihail Halatchev Le Gruenwald, "Estimating Missing Values in Related Sensor Data Streams" ADVANCES IN DATA MANAGEMENT 2005

[3] Le Gruenwald, Hamed Chok, Mazen Aboukhamis, "Using Data Mining to Estimate Missing Sensor Data" In 2007 IEEE

[4] Anjan Das, "An Enhanced Data Reduction Mechanism to Gather Data for Mining Sensor Association Rules" In 2011 IEEE