© 2016 IJSRSET | Volume 2 | Issue 3 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

A Research on Prediction of Missing Sensor Data Using Association Rule

Hitarth Chauhan, Prof. Bakul Panchal
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

A. Motivation

Recently there are many data prediction techniques available for predicting streaming data. Those techniques are known as avoidance techniques which prevent from storing missing data in the system. Techniques like WARM (Window Association Rule Mining), AKE (Applying K-nearest Neighbor Estimation) are useful in this process. Many other statistical estimation techniques are also available to predict missing data. All these avoidance techniques saves us form feature treble but we need some techniques to estimate data in previously fetched datasets. We can estimate missing value in such incomplete databases while retrieving data from it. We will use some association rule mining techniques like warm with some modifications to predict the data batter way.

B. Objective

- Gather sensor node data set
- Study data estimation techniques
- Analyze freshness component in gathered sensor data
- Try to use freshness factor in association rule
- Compare estimated result with original results to estimate performance.

C. Scope

1. Existing System

- Many avoidance strategies exists for data estimation in sensor data stream like WARM and K – Nearest Neighbor.

- We also need to predict the missing values in existing data set.
- These association techniques only consider frequency to find relation between two sensor nodes they do not consider at what duration this association existed.

2. Proposed System

In proposed technique we uses freshness factor in finding association rule. Rather than simply finding association on the bases of frequency of pair we will use weighted association and will assign more weight to the more fresh data. The proposed technique will improve the quality of the prediction.

D. Proposed Methodology

As we show conclusion of all these papers, missing data is one major problem in senor data, which can be caused because of many different reasons. But to use such incomplete data for further analysis can lead us to wrong conclusion so we need to find those missing values first. And to do so we have seen many approaches.

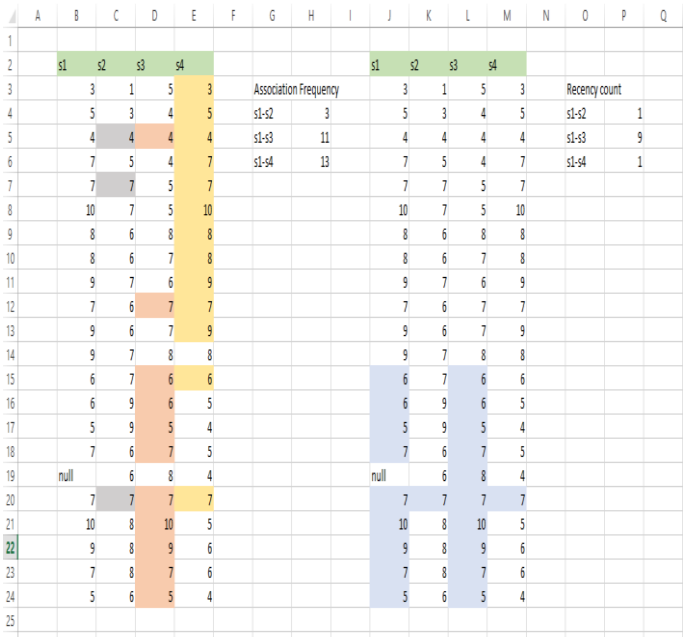
This process of estimation of missing data can be done while revising data stream from sensors or after storing data in the database. Currently there exist many approaches to estimate missing data form data stream. But not for the sensor data that has been already stored and gathered [1]. We have studied one approach in first paper.

We have seen that we can use WARM and max-WARM to predict missing data in incomplete database. Using this approach we can predict missing data while querying incomplete database. In this approach we simply creates association rules for each pair of sensors and one which will satisfy both minSupport and minConfidence that pair will create one rule.

I would try to improve quality of the estimated data. In process of estimation we can include recency of data to improve quality of estimation. Value of current sensor node is always influenced by value of previous round value. I will include process one more step in estimation process in which I will try to find nodes which are more recently related to the missing sensor node.

Proposed work flow

1. Find the result of the requested query form the existing database.
2. If result do not contain missing value than we can simply return results to the user and stop further process else proceed to the next step.
3. If we find missing value in result then calculate all paired association having missing sensor node as one part of pair.
4. With all association we find we will check recency association status.
5. Sensor which is more recently associated with missing sensor node will be finalized and value of that node in same time frame will be used to put at missing value place.



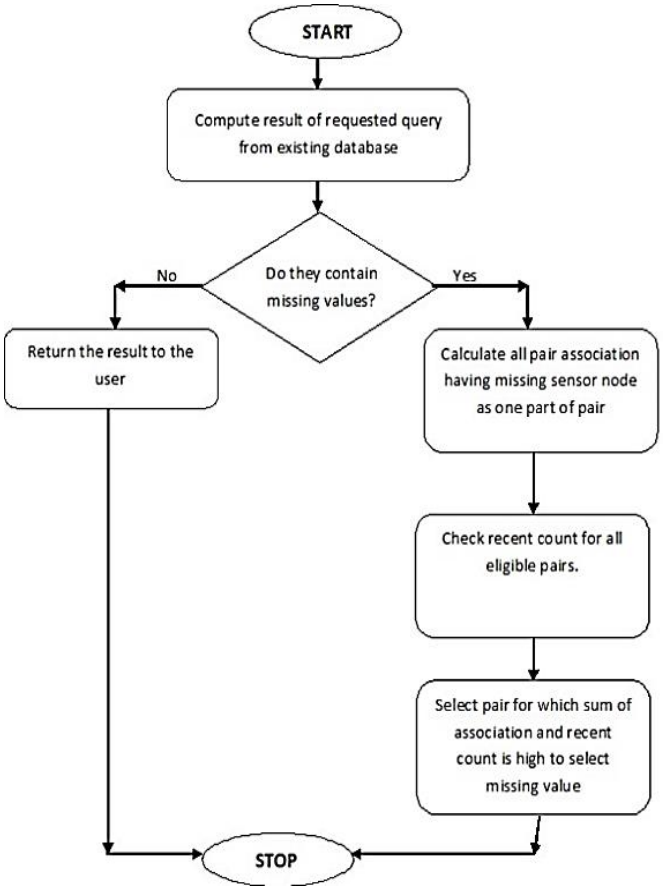
	s1	s2	s3	s4
	3	1	5	3
	5	3	4	5
	4	4	4	4
	7	5	4	7
	7	7	5	7
	10	7	5	10
	8	6	8	8
	8	6	7	8
	9	7	6	9
	7	6	7	7
	9	6	7	9
	9	7	8	8
	6	7	6	6
	6	9	6	5
	5	9	5	4
	7	6	7	5
	null		6	8	4
	7	7	7	7
	10	8	10	5
	9	8	9	6
	7	8	7	6
	5	6	5	4

Association	Frequency
s1-s2	3
s1-s3	11
s1-s4	13

Recency count	
s1-s2	1
s1-s3	9
s1-s4	1

Figure 1. Example

Flowchart



Proposed Data Structure

For generating only one or two frequent items we can use billow mentioned data structure. To store association between two sensors giving same result in given time slot we can use two dimensional jagged array. Each cell in this array will represent one pair of the sensors. If the cell contain value 1 then that pair contain the relation and if the pair contain result 0 then that pair do not have association.

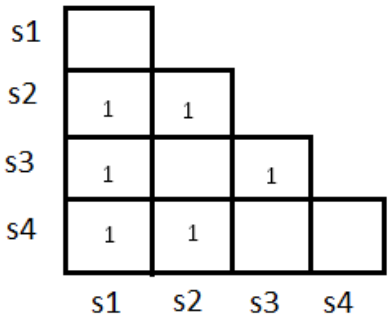


Figure 2. Database structure

As we can see in figure 1 we have association rule set for the pairs of sensor nodes. Hear pair s1s4, s1s3 and s1s2 are associated.

Proposed Algorithm

This algorithm tries to find recency factor between sensors of selected pair.

Input: Sensor id1, Sensor id2 , timestamp
Output: Recency factor R.

Step 1: Set R = 0; pTimestamp = timestamp - 1, fTimestamp = timestamp + 1
Step 2: while pTimestamp <= firstTimestamp repeat step 3
Step 3: If(valueS1[pTimestamp] = valueS2[pTimestamp]) then
R = R + 1; pTimestamp = pTimestamp - 1
Else back;
Step 4: while fTimestamp <= lastTimestamp repeat step 5
Step 5: If(valueS1[fTimestamp] = valueS2[fTimestamp]) then
R = R + 1; fTimestamp = fTimestamp + 1
Else back;
Step 6: return R;1

After calculating R value for all the eligible pairs sensor with highest value of R will be finalized and its value for given timestamp will be used for the place of missing value.

III. RESULTS AND DISCUSSION

IMPLEMENTATION

This Experiment is carried out on Intel CORE^2 Duo 2 GB RAM in 64 bit windows 10. Dataset for the experiment was taken form data.gov containing air temperature details. Data set was obtained on 15 March 2016, 12:54:42 PM. Tool used in implementation is Microsoft SQL Server 2012.

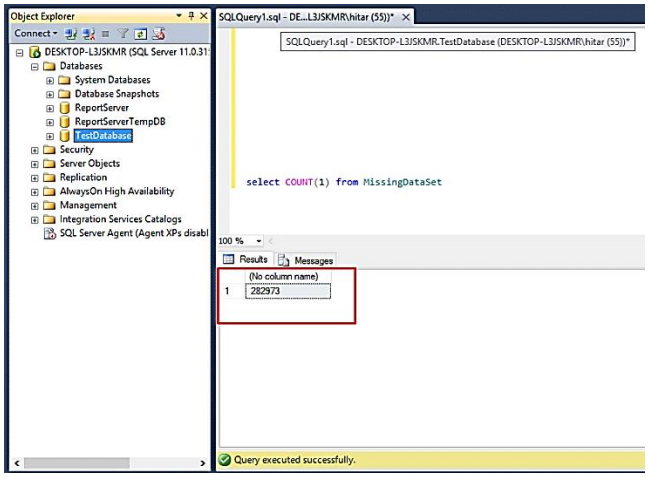


Figure 3. Total dataset

This image shows whole dataset that contains 282973 number of records. The experiment is being held on the very same dataset.

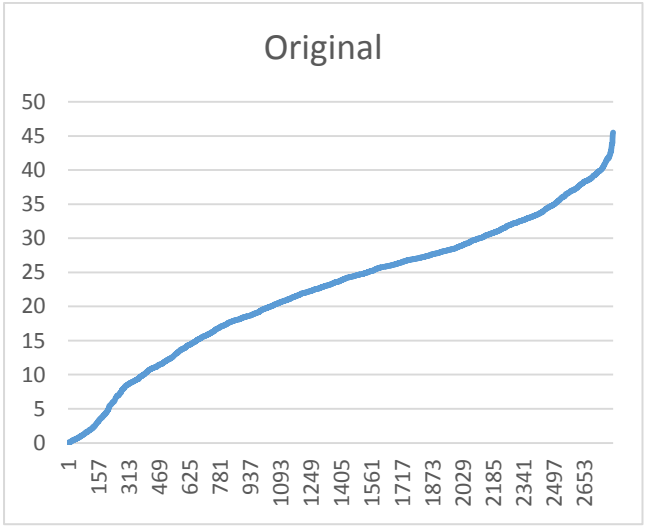


Figure 4. Graph of dataset deviation using existing approach

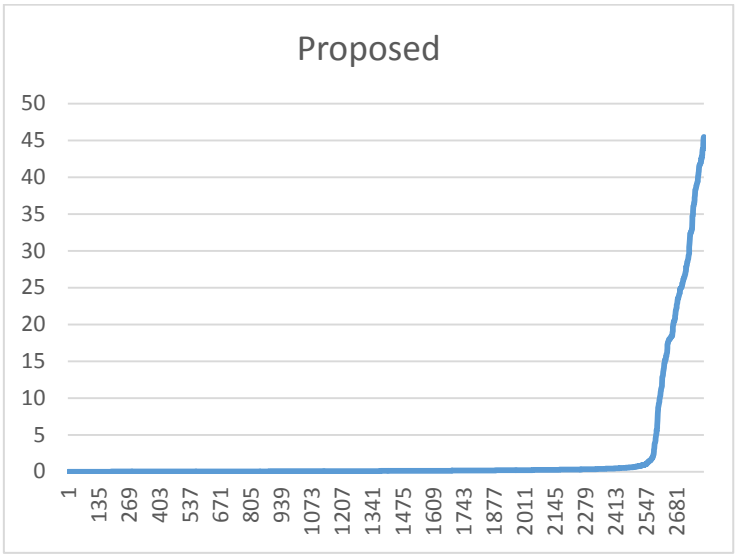


Figure 5. Graph of dataset deviation using proposed method

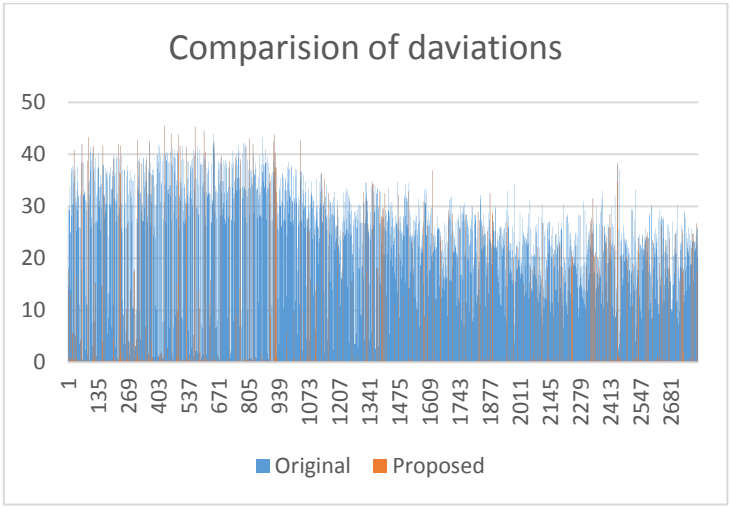


Figure 6. Comparison of deviations between existing method and proposed method

The chart showed above presents the comparison between the results of deviations of existing method showed in blue and proposed method showed in orange. Comparatively one can see that the amount of deviation decreases considerably.

IV. CONCLUSION

Approach we apply on the prediction technique will improve quality of the prediction. Considering recent data with high priority to decide association will lead us to batter estimation of missing sensor data and as final result output of requested query will be more accurate.

V. REFERENCES

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

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

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

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

[5] "Tutorials point",may 2014, http://tutorialspoint.com/

[6] https://www.techopedia.com/definition/30306/association-rule-mining

[7] https://en.wikipedia.org/wiki/Association_rule_learning.