Reducing the Time Factor for analyzing the Errors and Fixing in Big Sensor data

Anusha G, Dr. N Guruprasad
Dept. of Computer Science and Engineering, New Horizon College of Engineering, Bengaluru-560103.
anushagutta25@gmail.com, nguruprasad18@gmail.com

Abstract-- The big sensor data is so common in the industries and research application that grows so rapid that increase in volume and velocity is going to be meteoric. As the sensor data keeps on increasing it becomes a very difficult task to process using the on hand database tools and data processing applications. To overcome from the drawbacks the cloud computing has been provided with a unique feature of tackling with the challenges by providing the flexible massive storage, computing and software services in a lower cost. Many techniques has been developed for processing sensor data such as sensor cloud but it would not support fast error detection and location of error faster in big sensor data sets. In our proposed system real based approach for faster detection and location of errors in the sensor data sets. The error detection is based on scale free network topology which would dramatically accelerate the process. We fix the errors using the spatial and temporal correlation technique. This approach can reduce the time for detection‘s of the error and also location of the error in big data sets and main advantage is fixing the errors.

Keywords-- Big sensor data sets, Error detection, scale free network, types of errors, temporal correlation spatial correlation.

I. INTRODUCTION

The cause of data explosion in the present era the biggest challenge faces is processing of the big data. Since big data is collection of data sets and it so complex to process it as the data keeps on exploring. The traditional approach of human cognitive process which includes datasets which is beyond the ability to process the data in tolerable elapsed time which would be a major drawback since datasets keeps on accumulating day by day and becomes difficult task to process it. One of the major and important characteristic of big data is volume, velocity, value, veracity and variety. The big data sets can from any base such as meteorology, complex physics simulations, biological study and environmental research.

One important source of data set is collected by wireless sensor network (WSN). The WSN have feature of enhancing the ability of monitoring and interact with physical environment. Since there is corruption and lose of data due to presence of WSN in hardware inaccuracies in the node. It is necessary for data to be received clean and accurate. There is a need of effective detection and also cleaning of sensor big data is a major challenging and requires innovative solutions. WSN with the cloud can be called as complex network systems. As the complex network increases the data in accuracy and error has become an issue in real network application.

WSN big data error detection usually requires real time processing and also storage for massive sensor data which would also use the complex error model to detect the event of abnormality. In this paper we aim to develop a approach by having massive storage, scalability and also having computation power to detect error in big data sets from sensor data. The proposed error detection approach in this paper is by detecting the types of errors. The main work is to achieve time efficient approach in detecting the errors without compromising error detection accuracy and also the recovery of the error.

II. RELATED WORK

As there was increase in the data, the processing of the data sets also increased in era of data explosion. The different techniques for processing the data are as follows:

A) Big Data Processing on Cloud-- The cloud computing has a significant feature of big data storage and also interpreting it massive computation power. A design called "stream-as-you-go" is used for the processing the data has it is increased through the stream based management data architecture. Map Reduce is also used to process to analyze the incremental data and also process by dividing the data. But all this focus on the workload distribution, scalability and filtering of the data in speed but these approach is not enough for error detection and correction.

B) On- Cloud Processing for WSN -- The WSN are used in different fields such as in environment monitoring, military and also in scientific data collection. Since the variety of data has been collected from different fields through the WSN, there is a need of the sensor cloud for processing of the data. But due the nature of big data with the feature of volume, variety and velocity it is difficult to process the data in the complex network system. By effect of this using the cloud sensor it is difficult to develop the time efficient approach to detect the errors in the data sets and also to debug the complex network system in real time.

C) Data Error Detection in Sensor Network and Complex Network – The data error is unavoidable in real world complex network system. As there was dramatic increase in big data generated to locate the error was a quite challenging task with normal computing and network system.

Wang et al has proposed a classification of errors based on scenario analysis. It performed well in finding the errors detecting the errors. This work compares the robustness of four node network only, clustering co-efficient and centrality and hence can be extended for the complex network system.

Mukhopadhyay proposed error correction method for the WSN. It was performed on the intelligent sensor network. The intelligent sensor network correction of the error was faster. But as the big data sets increased its processing capability and time performance was extremely limited when the data was increasing.

In our approach we would detect the errors in the complex network by improve the time efficiency and faster detection and correction of the errors.

III. Proposed Work

In this paper we discuss about reducing the time for analyzing the errors and fixing of the error in the big data sets. We would use Time Efficient Error Detection with would consist of Scale Free Sensor Network System (SFSNS) which is used to process unlimited amount of data and Model based Error detection (MBED). Before applying this approach we need to classify the types of errors. The errors can be classified based on the numeric data and text file data. Since the errors are so common in complex network system. The architecture of the proposed system is follows which clearly specify the structure and flow of the system which is been shown in fig 1.

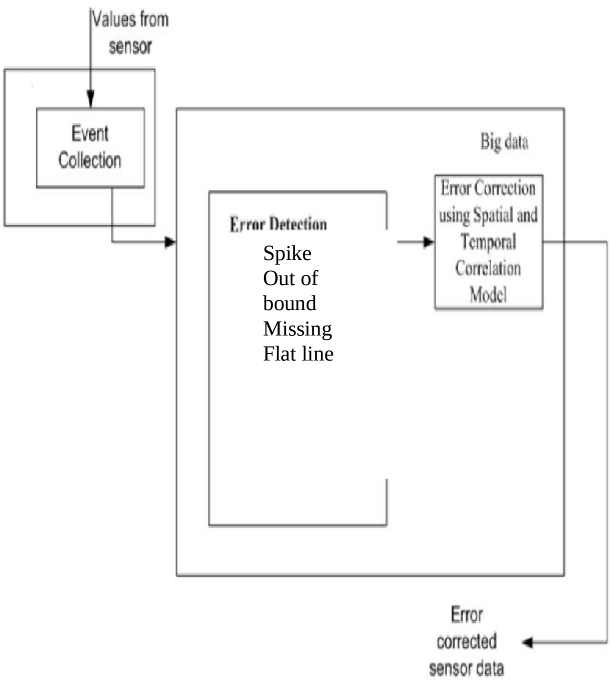

Fig-1 Architecture diagram

The errors are classified as follows

A) Types of Errors -- The flat line faults shown in fig-1 indicate that nodes in the network kept unchanged for unacceptable time series and for long duration. In real world application the transmitted data will have small changes over time flow.

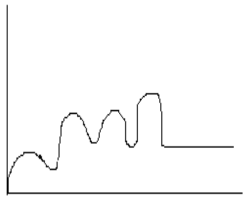

Fig-2 Flat line fault

The out of bound faults shown in fig-2, the value of the data can be observed based on the domain knowledge that is gained in general. In real world applications if a threshold is fixed for a wave, if any wave beyond that it is treated as out of bound error.

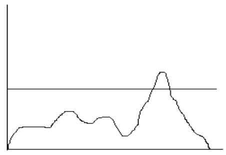

Fig-3 Out of bound fault

The data lost fault shown in fig-3, it means that there is missing of data over time during the data is generated and exchanged process which requires data cleaning.

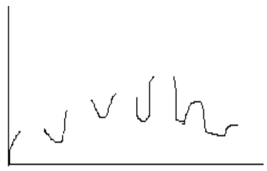

Fig-4 Data lost fault

The spike fault is shown in fig- 4 the spike error indicates that in a time series data items which would be out of prediction suddenly and normal over the time series.

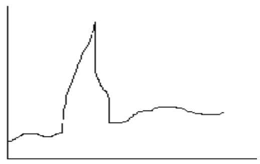

Fig-5 Spike fault

B) SFSNS -- The WSN graph has a strong feature of scale free complex network such as cluster head WSN. The scale free network consists in-homogeneous and also only few nodes in graph have large number of links in the nodes. In reality the cluster head WSN is almost similar to the scale free network. The scale free network has a more clustered hierarchical nodes topology. By getting the hierarchical structure of the graph we can easily traverse the nodes and detect the error.

C) MBED -- The complex networks systems have a similar to the clustered network topology. When there is a data abnormality in the big data datasets, there two tasks to be completed for detection the errors. There are to detected abnormal data is the true error that is "false negative" for detecting a true error and "false positive" for the non-error data. Finally detecting and tracking of the error source in which node. By these two tasks we can finalize successfully the error detection process. The error detection process needs to filter big data sets from the network. When there is data abnormality the whole network should be traverse for finalizing error source localization. In scale free network only few nodes in the hierarchy will have large set of links to the nodes. So based on node which has huge links can be grouped in the cluster so that the error can be located easily, so that we can need to navigate to search the error and the location of the source.

Finally for the error correction and for the recovery we use a different approach called spatial correlation and temporal correlations are used. In temporal correlation we can identify the sensors attack and the forged values are sent. In spatial correlation by having the knowledge of sensor we can correct the error easily. The proposed of clustering can reduce the time for detection of the error and also processing of the data of whole can be avoided.

IV. IMPLEMENTATION

In the implementation the sensor data is been collected through the simulation, the data sets through the simulation using the simulator. The sensor data is given to the pre processor to process the data. By using the error tracker I would track the errors. The algorithms used to track all four errors are mentioned below. The first used algorithm is for detecting the missing error algorithm.

Algorithm -1 Detection of missing error

Step 1 Collect the data sets that the id, time, value.
Step 2 Differentiate data sets for values based on ―#‖.
Step 3 Initialize the sensor id=0.
Step 4 for each i=0 to i++.
Step 5 Find the value of diff=time[0]-time[1].
Step 6 for each i=2 to i++.
Step 7 Find the value of dn=time[i]-time[i-1].
Step 8 If dn= diff , no error.
    Else missing error.

The detection of the flat line error algorithm is given in algorithm 2 below.

Algorithm-2 Detection of flat line error

Step 1 Initialize a variable time =0.
Step 2 Initialize obstime = -1.
Step 3 for each i=1 to i++.
Step 4 val= value[i]-value[i-1].
Step 5 if val=0.0
    time++
    obstime = 1
Step 6 Else time=0
    Obstime = -1
Step 7 if time is greater than 5
    Return flat line error

The detection of the spike error algorithm is given in algorithm 3 below.

Algorithm-3 Detection of spike error

Step 1 Get the value of the sensor.
Step 2 for each i=1 to i++
Step 3 Calculate prev=val[i-1]
    Current=val[i]
    next=val[i+1]
Step 4 if(current>prev &&current>next)
    Avg=(prev+next)/2
    Tim=Current/Avg
Step 5 if(tim>5)
    Return spike error
Step 4 Else no error

The detection of the out of bound error algorithm is given in algorithm 4 below

Algorithm-4 Detection of out of bound error

Step 1 Initialize the value of maxvalue=60.
Step 2 for I=0 to i++
Step 3 if value is greater than max value
    return out of bound error
Step 4 else return there is no out of bound error.

The entire above mentioned algorithm is implemented to detects the errors from the sensor data and finally corrected.

V. CONCLUSION

In the process of detecting the errors in the sensor network system, a new approach is been developed using the cloud computing. Initially the error classification in the sensor data sets is been differentiated. The correlation between sensor networks and the scale free networks is achieved. By using the correlation between both the error types and the scale free networks features a time efficient strategy for detecting the errors and also locating the errors in big data set is achieved. Finally the error detection approach can reduce the time for fast error detection in numeric big data sets and it achieve similar error selection ratio for non-scale free error detection. In accordance to the error detection of big data sets, it would also recovery correct the error and false positive ratio is also reduced.

REFERENCES

[1] I. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, ―A Survey on Sensor Network‖, IEEE Comm. Mag., vol. 40, no. 8, 2002.
[2] Wireless Sensor Networks, Technology, Protocols and Application by K. Sohraby, D. Minoli and T. Znati.
[3] S. Mukhopadhyay, D. Panigrahi, and S. Dey, ―Data Aware, Low Cost Error Correction for Wireless Sensor Network,‖ proc. IEEE Wireless Comm. And Networking Conf.
[4] ―Sensor Cloud,‖ http://www.sensorcloud.com.
[5] S. Sakr, A.Liu, D. Batista, and M. Alomari, ―A Survey of Large Scale Data Management

Approaches in Cloud Environment,‖ IEEE Comm. Survey.

[6] M.C. Vuran and I. Akyildiz, ―Spatial Correlation based on Collaborative Medium Access Control in Wireless Sensor Network‖

[7] W. Dou, X. Zhang, J. Liu, and J. Chen, ―HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big DataApplications,‖ IEEE Trans. Parallel and Distributed Systems, 2013.

[8] Armbrust, A. Fox, R. Griffith, A.D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, ―A View of Cloud Computing,‖ Comm. the ACM, vol. 53, no. 4,pp. 50-58, 2010.

[9] R. Buyya, C.S. Yeo, S. Venugopal, J. Broberg, and I. Brandic, ―Cloud Computing and Emerging it Platforms: Vision, Hype, and Reality for Delivering Computing As the 5th Utility,‖ Future Generation Computer Systems, vol. 25, no. 6, pp. 599-616, 2009.

[10] L. Wang, J. Zhan, W. Shi, and Y. Liang, ―In Cloud, Can Scientific Communities Benefit from the Economies of Scale?‖ IEEE Trans. Parallel and Distributed Systems, vol. 23,, no. 2, pp. 296-303, Feb.