International Refereed Journal of Engineering and Science (IRJES)
ISSN (Online) 2319-183X, (Print) 2319-1821
Volume 2, Issue 6 (June 2013), PP. 16-22
www.irjes.com

Sensor Data Mining Model and System Design: A Review

Chhaya Dule[1], K.M.Rajasekharaiah[2]
[1]*CSE Dept., Jyothy Institute Of Technology Bangalore, INDIA.*
[2]*CSE Dept, Jnana Vikas Institute of Technology Bangalore, INDIA.*

ABSTRACT: *The sensor data, which is inputted from sensor Network, is stream data having continuous and infinite properties. The previous data mining techniques cannot be used on sensor data because of these properties. Application services in the sensor network are only event alert services which perceive the events from sensors and alert the events to the supervisor. In this paper, we define continuous sensor data mining model and design a system based on the model. The system can service useful knowledge by continuous sensor data mining using gathered data from sensor in the sensor network. Sensor data is classified into three data types, which are simple sensor data, continuous sensor data, and sensor event data, and define sensor data mining models about outlier analysis, pattern analysis, and prediction analysis. After the definition of model, we design a system which can be used in application services City Environment Management, etc., based on these mining models in sensor network environment.*

Keywords:- *Sensor Data Mining , Stream Data*

I. INTRODUCTION

Recently, everything like computer, electronics, automobile living machinery, medical machinery, transportation systems, plants and animals, etc., has been connecting each other through ubiquitous network in electronic and physical space. We are able to gather various physical phenomenon through the sensors from the environment that requires continuous monitoring and that is difficult to access[5,6,7,8].Recently, the massive and continuous sensing data can be gathered by real-time through sensor network or home network because of the development of the wired-wireless communication system and sensing technology. But analysis of data and extraction of knowledge for the sensor data depends on the process of the stream data mining method and traditional data mining method[2,3,4].The studies ,which focus on the sensor data ,are level to perceive some events and to alert the events[5,7,9].

Generally, sensing data is classified like simple sensor data, continuous sensor data, and sensor event data. The simple sensor data indicates numeric value which is sensed by periodic or request. The continuous sensor data indicates signal value which is sensed continuously. The continuous sensor data is classified according to two types; one is sensing data during specific time interval and the other is summarized sensing data because we can't store the whole sensing data[1]. The sensor event dataindicates generated value when the data is overthreshold value into the sensing data.Because of these continuous and infinite properties,previous data mining techniques can't be used tosensor data mining immediately. Therefore, weshould use data mining techniques according tonewlydeveloped method or transformed method of theprevious data mining techniques. Also, we needsensor data mining model which can search usefulknowledge through mining of the sensor data and asystem which processes sensor data like stream databuffering, retrieval, and storing.

In this paper, we define continuous sensor datamining model which can search the useful knowledgebased on gathered information from sensors in theubiquitous sensor network(USN) environment anddesign a system which includes the models. For themining model and system design, first, we classifiedthe sensor data to the three data types like simplesensor data, continuous sensor data, and sensor eventdata. And then we did define the each mining modelsabout outlier analysis, pattern analysis, and predictionanalysis according to the three data types. Finally, wedid design a continuous sensor data mining systemwhich embeds the defined sensor data mining models.The rest of paper is organized as follows. In Section II, we define the continuous sensor data mining modelfor the above described traditional mining techniques.And in Section III, we design a sensor data mining system. Finally, in Section IV we summarize our study.

II. SENSOR DATA MINING MODEL

In this section, we first define various definitions for sensor data mining model and then discuss sensor data mining model for the sensor data.
Let denote the sensor S and sensor node SN. The sensor is defined like following Definition.

2. 1 Definition Of Sensor

[Definition 1]*S= {s | s Є sensor}, SN= {n | n Є sensorNode}*

For the sensor data mining, we define the sensor class to select the sensors of specific type like temperature sensor, humidity sensor, etc. The sensor class is consisted of location and sensing type. Let denote sensing type ST. The ST is defined like following Definition 2.

[Definition 2]*ST = {t | t Є sensingTypeSpec}*
Let denote a specific sensor type S_t. the S_t is defined like following Definition 3.

[Definition 3]*St = {s | s Є S∩ST ^ t.sensingTypeName={a,b, ...}, t Є ST}*
Let denote sensing area S_l. The S_l is defined like following Definition 4.

[Definition 4] S_l *= {s | s Є S∩SN ^ l.sensorNodeLocation ={a, b, ...}, l Є SN}*
Let denote a specific sensing area and sensor type $S_{l,t}$. The $S_{l,t}$ is defined like Definition 5.

[Definition 5]*S l .t = {s | s□ Є S l ∩St}*
Time interval specifies specific time point recorded in sensor database or specific time point which will be recorded in the future. Let be denote the time interval T. The T is defined like Definition 6.

0, snapshot
[Definition 6]
a>a,0 // Time window

2. 2 Outlier Analysis

The outlier analysis is to extract the abnormal sensor value from sensor data. If user defines the specific time point recorded in sensor database for the sensor data or the future specific point and then present the range of normal value then outlier value is extracted. The range of normal value is expressed by threshold or probability. We call the outlier analysis technique TOA (Technique of Outlier Analysis). TOAis defined like Definition 7.

[Definition 7]*Y= {r | r Є TOA(s, t , ts, rt) ^r Є R0 , s Є S l ,t , t Є T , rt Є RT}*

Here, we define four variables which are s, t, ts, and rt. The "s" is sensor class which expresses the range of the sensor. The "t" is time interval which expresses time window. The "ts" is threshold. The "rt" is rule type. We already defined "s" and "t" in Definition 1-6. The rest variables, "ts" and "rt", are defined in Definition 8, 9.
The threshold in the outlier analysis is specified confidence value for the estimated value. The outlier is extracted based on the confidence value. We denote the variable "ts" TS.

[Definition 8]*TS= (*TS_C *,* TS_S *)={TSc= a* TS_S*=b,*0<= a<= 1,0<=b<=1}

The rule type is to select a outlier technique of the outlier techniques in the sensor data mining engine. The outlier analysis is classified by general outlier (GO), spatial outlier(SO), and continuous outlier(CO). Therefore, the rule type is defined like Definition 9.

[Definition 9]*RT□ □{r | r□ □{GO, SO, CO}}*

Each rule format for the outlier analysis is classified as follows. The GO format is (value, confidence, time interval),time interval or until the specific time point.the SO format is (value, location, confidence, time interval), and the CO format is (value, time interval, confidence). We can see the definition for each format in the Definition 10.

[Definition 10]

RT=GO= {(v, c, t), v Є R, c>=TS, t Є T}

R0= ***RT=SO= {(v, l, c, t), vЄR, lЄSl, c>=TS,t ЄT}***

RT=CO= {(v, t,c), v Є R, t Є T ,c>=TS}

2. 2.1 General Outlier Analysis

The mining model for the general outlier analysis is shown in Fig. 1.

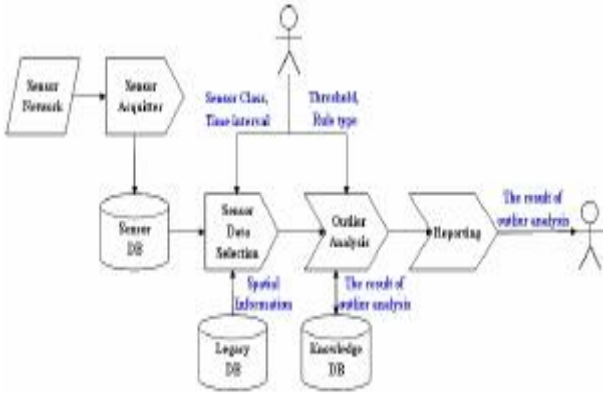

Fig 1.General Outlier Analysis Model

In the general outlier analysis model, the sensor data, which is included in the sensor class of specific time points inputted from the sensor database, is selected through the sensor data selection process using the sensor class and time interval which are inputted by user. This time, the spatial information,Which includes location information of the real sensor data, is also inputted from legacy database? Next, the outlier analysis is conducted using the selected sensor data, inputted threshold, and rule type. As the result, the outlier, which has the form of (outlier sensor value, threshold) or (outlier sensor value, probability range), is extracted. The outliers after the mining are stored in the knowledge database and sent to the user. Thestored outliers are hereafter used through the knowledge database when user conducts the query of same condition.

2.2.2 Continuous Outlier Analysis

The continuous outlier analysis is to extract continuously the outliers for general outlier or spatial outlier analysis from continuous time window which is defined during the specific time interval or until the specific time point.

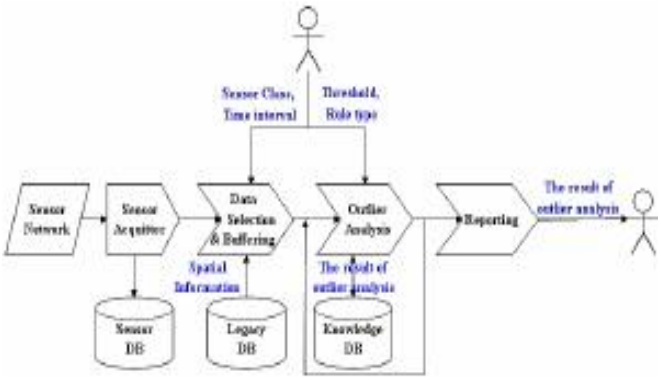

Fig 2 Continuous Outlier Analysis Model

In the sensor data model, the data pointers can access the data according to entered order of the sensor data but the random access is impossible .The space of data store is very small as compared with massive sensor stream data. Therefore, the mining for the sensor stream data should be conducted through once data scan. Also, the clusters become the outliers according to the stream of time and the data distribution is changed by random for the sensor stream data. Therefore, the outlier analysis is verydifficult. The mining model for the continuous outlieranalysis is as the fig 2.

The continuous outlier analysis should be always buffering the sensor data in contrast to sensor data input of the specific time point from sensor database. So, if the new sensor value is inputted, the past sensor value should be deleted. Like this, although the current sensor value is normal value, the sensor value can be changed to abnormal value later. Therefore, the current outlier and the other outlier, which can be changed to outlier later, are selected and kept up.

2.3 Pattern Analysis

The pattern analysis is to search the pattern like trend, cyclic for the gathered sensor data. The trend pattern is a description for the generalization and summarization of sensor data. The cyclic pattern is a description for the sensor data which appears repeatedly during the time interval. We call the patternanalysis

technique TPA(Technique of Pattern Analysis) and are defined by Definition 11.

[Definition 11] $Y = \{r \mid r \in TPA(s, t, ts, rt) \wedge r \in R_P, s \in S_{1,t}, t \in T, ts \in TS, rt \in RT\}$

Here, we also define four variables which are s, t, ts, and rt. The variables, "s", "t", and "ts", were defined in Section II.2. The rule type, "rt", is defined differently. The rule type is temporal pattern (TP), spatial pattern (SP), and continuous pattern(CP). The variable "rt" is defined in Definition 12.

[Definition 12] $RT = \{r \mid r \in \{TP, SP, CP\}\}$

Each rule format for the pattern analysis is classified as follows. The TP format is (pattern, calendar time, support, confidence), the SP format is (pattern, location, support, confidence), and the C format is (pattern, time_interval, support, confidence).We can see the definition for each format in the Definition 13.

[Definition 13]

$$RT=TP=\{a \quad b \quad d\ldots.c), a, b, d \in I, c >= TS, t1, t2 \in T\}$$

(t1 t2)

$$Rp= \quad RT=SP=\{a \quad b \quad d\ldots.c), a, b, d \in S, c >= TS, t1, t2 \in T\}$$

$$RT=CT=\{a \quad b \quad d\ldots t,.c), a, b, d \in I, c >= TS, t \in T\}$$

2.3.1 Temporal PatternAnalysis

The Mining model of Temporal pattern analysis is shown as given below

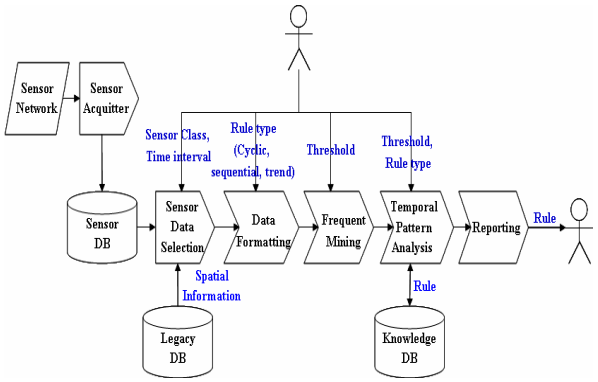

Fig.3 Temporal Pattern Analysis Model

The Temporal pattern analysis is classified by cyclic pattern, trend pattern and sequential pattern. These pattern analysis conduct the frequent pattern mining and then extract temporal patterns which satisfy the user defined support threshold so, The model conducts the process of frequent mining in common. The data formatting process defines the pattern type of the cyclic, trend and sequential .The result of mining has temporal pattern, support and confidence.

2.3.2 Spatial Pattern Analysis

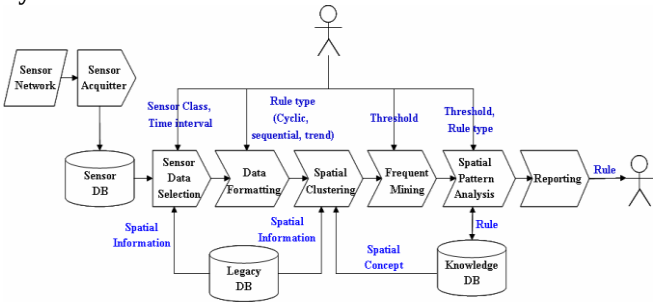

Fig4 Spatial Pattern Analysis Model

The Spatial pattern analysis is to analyze consideration of sensor value for snapshot and spatial attribute of the corresponding sensor. The spatial pattern analysis mainly uses spatial attributes and extracts similarity with the neighborhood density and locality. If the user specifies the snapshot for the sensor value and pattern type then the spatial pattern for snapshot is returned. The spatial pattern analysis differently with the temporal pattern analysis conducts spatial clustering for the sensor data before the pattern mining.The spatial frequent pattern search isconducted in consideration of spatial concept based on spatial clusters and then spatial pattern analysis is conducted based on the frequent pattern.

2. 4 Prediction Analysis

Prediction analysis continuously extracts the similar pattern during the specific time interval or until specific time point using the past temporal pattern this technique called Technique Of Prediction Analysis Based On Pattern(TPAP).

[Definition 14]Y={*r* | *r*□ □*∈ TPAP(s, t, ts, rt)* ^□ □*r*□ *∈ R_P,s ∈ S _{l,t},, t*□ *∈ T , ts*□ *∈ TS ,rt*□ *∈ RT* }

Four variables s, t,ts and rt .The variables s,t and ts were defined in Section 2.2.The rule type "rt" is continuous prediction analysis (P) and shown in definition 15

[Definition 15]Y={*r* | *r*□ □{P}}

For prediction analysis, prediction function is needed .prediction function is defined as V=f(t).The prediction analysis format is (predictor,time_interval,confidence) based on the prediction function. The prediction Analysis format is shown in the definition 16

[Definition 16]Rp={*r* | *r*□ □*∈ (f(t),t,*ts) , *t*□ *∈ T , ts*□ *∈ TS* }

2. 4.1 Continuous Pattern Analysis

The continuous pattern analysis model is shown below

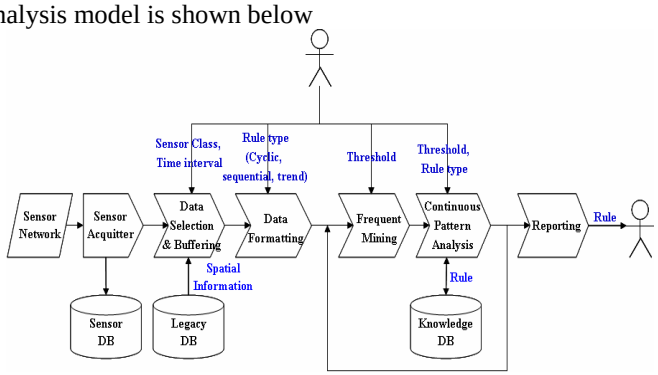

Fig5 Continuous Pattern Analysis Model

The continuous pattern analysis analyzes patterns for the data within the sliding window through buffering of sensor data. In sliding window, though a pattern is frequent pattern within current sliding window, after sliding of the window, the pattern can be non frequent pattern. In contrast though a pattern was not frequent pattern, the pattern can be frequent pattern. therefore, the important thing in continuous pattern analysis is to manage the potential frequent patterns and to prune patterns which will be non frequent pattern.so it is needed to develop algorithm of potential frequent pattern search and non frequent pattern pruning.

2.4.2 Prediction Analysis Based On The Pattern

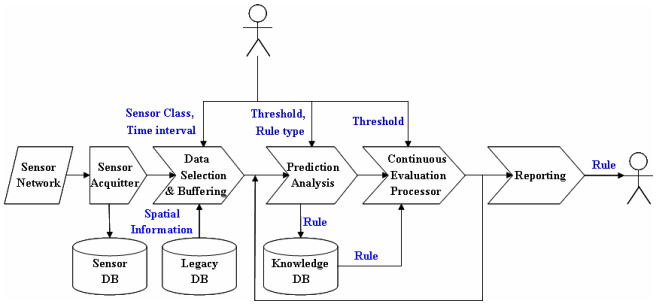

Fig6 Pattern Based Prediction Analysis Model

The pattern based prediction analysis is to process prediction query to find similarity pattern in the sensor data using previously searched temporal pattern. The prediction query is a process to find the prediction value of specific time point and point of specific value. The model for pattern based analysis model is shown as above.

The most important thing in pattern based analysis model is the continuous estimation process for the prediction model .First; it makes the prediction model for the sensor data within sliding window through prediction analysis and then prediction model stored in Knowledge DB.The past prediction model can be incorrect in the current sensor data because of sliding of window therefore the estimation isprocessed between the current sensing data and the prediction model. If the error is more than the user defined threshold, the prediction model is updated because model is not useful.so the efficient development of prediction estimation model is the most important in the pattern based prediction analysis.

III. DESIGN OF SENSORDATA MINING SYSTEM

In this paper ,we will discuss a system SDMC (Sensor Data Mining Component) .the system can service useful knowledge through the continuous sensor data mining based on sensing data from sensor network. The system structure diagram shown below.

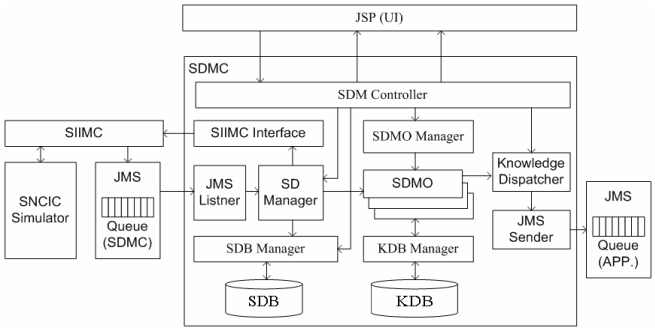

Fig7 Sensor Data Mining System Structure Diagram

3.1 JSP UI (User Interface) Module

In the UI model ,The sensor for mining sensor data and data mining technique is selected ,parameters are selected for applying data mining technique and search results are displayed.

3.2 SDM Controller Module

The SDM (Sensor Data Mining) controller is the core module which launch and manage the whole module .If the system is launch ,The user interface launch SDM controller and SDM initialize whole modules and send the information for the modules. If user specifies parameters for sensors and sensor data mining. The SDM launch each module according to the user specific parameters and then send result in JSP UI.

3.3 SD Manager Module

The Sensor Data Management module (SD Manager) furnishes the abstraction for the real sensor data. The SD Manager has the sensor data queue. So SD manager stores the sensor data which is sent from SIIMC in the queue. therefore the system furnishes the convenience because it processes the data to the level of abstract data.

3.4 SIIMC Interface Module

The SIIMC and SNCIC is java based upon libraries and conduct transmission of sensor data from sensor network.The SNCIC is USN middleware component which furnishes the abstraction to acquire sensing data from sensor network. It also manages sensor query and sensor data through SIIMC and SNCIC the sensor data is acquired and managed.

The function of SIIMC interface is to transmit sensor data to SDMC .To send and receive the sensor data XML document is used. If sensor data is requested to the SIIMC .It transmits the sensor data to Open JMS Queue. The data is stored in the “siimcqueue”. The JMS listener which is state of ready as a listener, receives from the siimcqueue and inserts data to the “sdmqueue” after conversion of each abstract data type ,data sent to SD Manager as described in the SD Manager module.

3.5 SDB Manager Module

The sensor database management module (SDB Manager) furnishes the abstraction for sensor database. That is ,we define class for each schema tables and records and abstract to do support in everything database,

The data operations like insert, delete and update.Then,the data can be used in rest module and the ratio of source reuse can be enhanced.

3.6 SDMO Manager Module

The sensor data mining operator management module(SDMO Manager)manages the mining technique like outlier analysis ,pattern analysis and prediction analysis, The mining technique are furnished by the SDMC. The mining and parameters selected by user through the JSP UI are transmitted to the SDMO Manager through the SDM Controller. This module launches the corresponding mining module using information like mining technique and parameters. After this, the module receives the result of mining and then sends to the JSP UI.

3.7 SDMO Module

The sensor data mining operation module(SDMO) launches a sensor data miming of the outlier analysis, pattern analysis and prediction analysis to search the knowledge. The SDMO returns the searched knowledge to the JSP UI through the SDMO Manager and stores the knowledge in the knowledge database(KDB) through the KDB Manager.

3.8 KDB Manager Module

The knowledge database Management module(KDB Manager) furnishes the abstraction for the KDB.The mean is to be able to use the knowledge data in everything database through abstraction of operations like insert,delete and update.

3.9 KDB Dispatcher Module

This module does monitoring the inputted real time sensor data based on the searched knowledge through the SDMO. Then, the JMS Sender transmits the corresponding information to the user.

IV. CONCLUSION

The sensor data model for the sensor data mining is based on the sensing data from USN environment .A Sensor data Mining system SDMC is designed based on the model. This model service the useful knowledge according to continuous query based on gathered information from sensor.This model furnish intelligent service in City Environment Management. The sensor data Mining system not only support event alert service but also service based on useful knowledge.In future we can implement the sensor data mining techniques which can be operated on this model.

REFERENCES

[1]. H. Han, H. Ryoo, and H. Patrick, An Infrastructure ofStream Data Mining, Fusion and Management for Monitored Patients, In Proceedings of 19th IEEE International Symposium on CBMS, 2006, pp.461-468.

[2]. Y. Peng, G. Kou, Y. Shi, and Z. Chen, A SystemicFramework for the Field of Data Mining and KnowledgeDiscovery, In Proceedings of Workshop on The 6th IEEEInternational Conference on Data Mining, 2006.

[3]. M. M. Gaber, A. Zaslavsky, and S. Krishnaswamy,Mining Data Streams: a Review, In ACM SIGMOD, Vol. 34, No. 2, 2005, pp.18-26.

[4]. L. Golab and M. T. Ozsu, Issues in Data StreamManagement, In SIGMOD Record, Vol. 32, 2003

[5]. B. Babcock, S. Babu, and M. Datar, et al., Models andIssues in Data Stream Systems, Invited Paper in Proceedingsof the PODS, 2002.