A Study on Sensor Data Mining Model and System Design

T. P. Abhinayaa Devi, N. Ramya*, R. Hemalatha
Department of Computer Science, Tiruppur Kumaran College for women, Tiruppur, TN, India
*nramyainfotech@gmail.com

ABSTRACT

The sensor data which is inputted from sensor network is a stream data having continuous and infinite properties and because of these properties the previous data mining techniques cannot be used on sensor data. Application services in the sensor network are only event alert services which distinguish the events from sensors and ale rt to the supervisor. In this paper, we define continuous sensor data mining model and design. The system can extract useful knowledge in continuous sensor data mining using gathered data from sensor in the sensor network. Sensor data is categorized into three data types, which are simple sensor data, sensor event data and continuous sensor data. The sensor data mining models describe and define about the outlier analysis, pattern analysis, and prediction analysis. After the definition, we design a system which can be based on the mining models in sensor network environment.

Keywords: Sensor data; Sensor data mining; Stream data; Stream processing.

1. INTRODUCTION

Recently, the massive and continuous sensing data can be collected by real-time through sensor network because of the development of the wired-wireless communication system and sensing technology. Sensors produce large volumes of data continuously over time, and this leads to several computational challenges. Such challenges arise both from accuracy and scalability perspectives. But analysis of data and extraction of knowledge depends on the process of the stream data mining method and traditional data mining method (Peng *et al.* 2006; Aggarwal and Yu, 2007; Golab and Ozsu, 2003).

1.1 Sensor Sata - Classification

The sensing data is classified into simple sensor data, continuous sensor data, and sensor event data. The simple sensor data denotes numeric value which is sensed by periodic or request. The continuous sensor data denotes signal value which is sensed continuously. The continuous sensor data is classified according to two types; first is sensing data during specific time interval and the another is summarized sensing data because we can't store the whole sensing data (Aggarwal, 2009). The sensor event data denotes generated value when the data is over threshold value into the sensing data. We should use data mining techniques according to newly developed method or transformed method of the previous data mining techniques. For the mining model and system design,

first, we classified the sensor data to the three data types like simple sensor data, continuous sensor data, and sensor event data. And then we did define the each mining models about outlier analysis, pattern analysis, and prediction analysis according to the three data types. Finally, we did design a continuous sensor data mining system which embeds the defined sensor data mining models.

2. SENSOR DATA MINING MODEL

In this section, we first define various definitions for sensor data mining model.

2.1 Outlier Analysis

The outlier analysis is to extract the abnormal sensor value from sensor data. If user defines the exact time point recorded in sensor database for the sensor data and then presents the range value then outlier value is extracted. The range is expressed by threshold or probability. There are two types of outlier are listed below.

General Outlier Analysis

In the general outlier analysis, the sensor data, which is added in the sensor class of specific time points inputted from the sensor database, is choose through the sensor data selection process using the sensor class and time interval which are inputted by user.

Continuous Outlier Analysis

The continuous outlier analysis is extract continuous data which is defined during the specific time interval or until the specific time point.

2.2 Pattern Analysis

The pattern analysis is to search the trend and cyclic pattern for the gathered sensor data. The trend pattern describes the generalization and summarization of sensor data. The cyclic pattern describes the sensor data which appears repeatedly during the time interval.

2.3 Prediction Analysis

Prediction analysis continuously extracts the related pattern during the specific time interval or until specific time point using the past temporal pattern. This technique is called as Technique of Prediction Analysis based on Pattern (TPAP).

3. MINING SENSOR DATA

The different data mining methods such as clustering, classification, frequent pattern mining, and outlier detection are often applied mine the sensor data. This data usually extract and filter the information. The conventional mining algorithms are not designed for real time processing of the data which is the main challenges. Therefore, new algorithms for sensor data need to perform the analytics in a single pass in real time. The problems of stream compression (Aggarwal and Yu, 2007) and stream mining are therefore tightly integrated together from an efficiency perspective.

- Clustering is the task of grouping data objects. The member of a cluster groups a similar as well as different possible data.
- Characterization is the task to acquire a compact description for a selected subset of objects.
- Classification refers to the task of finding out a set of classification rules that determine the class of any object forms the values of its attributes.

We need to develop an effective method for determining spatial and non-spatial relationships between datasets – i.e. data mining and knowledge discovery.

4. SENSOR DATA MINING SYSTEM ARCHITECTURE

In this paper, SDMS (Sensor Data Mining System), can examine the useful knowledge through the continuous sensor data mining based on sensing data from sensor network. Figure 1 represents the SDMS (Hansen *et al.* 2006).

Developing a simple data infrastructure requires some degree of standardization among the various data sets. The initiative on architecture, standards and metadata will act as the main guidelines in this task. An overall frame for the data infrastructure including Web-based services that enable participant to discover and download appropriate data for their work will be designed and a prototype will be developed. Standard (off-the- shelf) GIS software will be applied for analysis, modeling and visualization purposes at the client side.
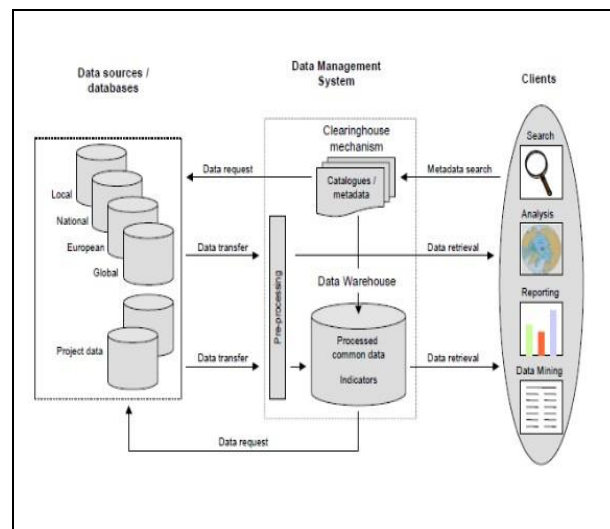

Fig. 1: Sensor data mining system

The main aim of the SDMS will be to support handle the data. To do this the system will include the following components

Data Warehouse
    Geoportal (Clearinghouse mechanism)
    Metadata reporting system
    Upload and download of data
    Pre- and post processing tools

DATA WAREHOUSE

In general a Data Warehouse is a large database organizing data from various sources in a repository assisting query and analysis. The database is well designed and contains key data which is importance for the organization.

GEOPORTAL

The term Geoportal is searching for data using geographic location, time and thematic attributes, has nearly replaced the earlier term data clearinghouse. Further developing, a Geoportal will be a web site that

represents an entry point to sites with geographic content.

METADATA

Efficient use of geographic information assuming access to documentation that describes origin, age, ownership and fitness for purpose. This associated information is referred to as metadata. Metadata is data about the data.

UPLOAD AND DOWNLOAD OF DATA

In the sensor, the server side can upload the data from the data warehouse. And the client can download the data what the user requested.

PRE-AND POST PROCESSING TOOLS

The interactive tool which provide GUI interaction to the end user. For PRE-processing the request method GET() method is used. For POST-processing the retrieve method PUT() method is used.

The objective under work based on data protocols and system requirements which works and developed based on GIS. In the client sides, the data can be searched first after that it pass through the data management system catalogues to perform the clearing mechanism, after that data is requested to the remote sever database, the data are searched in sensor data and it is preprocessed before it is transferring to the data management system, and the data is received to the client using the data mining techniques

5. SENSOR TECHNIQUES

In recent years, there has been large growth in the data generated by sensor networks. The most important category of techniques is model-based techniques. These techniques use mathematical models. It is used for solving various problems concerning to sensor data acquisition and management (Aggarwal and Han, 2014). Model-based techniques use different types of models: regression-based, machine learning, statistical, signal processing, probabilistic or time series. The work concentrate in the following four broad categories of sensor data management tasks are data acquisition, data cleaning, query processing, and data compression.

6. A SURVEY OF MODEL-BASED SENSOR

DATA ACQUISITION AND MANAGEMENT

Data acquisition
Data cleaning
Query processing
Data compression

A. DATA ACQUISITION

Sensor data acquisition is the task responsible for efficiently gaining samples from the sensors in a sensor network (Saket Sathe *et al.* 2006). In the literature, there are two major types of acquisition approaches: pull-based and push-based. In the pull based approach, data is only acquired at a user-defined frequency of acquisition. In the push-based approach, the sensors and the base station agree on an expected behavior; sensors send data to the base station if the sensor values deviate from such expected behavior.

B. DATA CLEANING

Data Cleaning is a process of removing the noise and inconsistent data. The data obtained from the sensors is often erroneous (Saket Sathe *et al.* 2006). Erroneous sensor values are mainly generated due to the following reasons: (a) irregular loss of communication with the sensor, (b) sensor's battery is ejected; (c) weather change etc.

C. QUERY PROCESSING

Processing queries is another important factor in sensor data management. Main objectives of these techniques are to process queries by accessing data. Model based techniques that access/generate minimal data and also handle missing values (Saket Sathe *et al.* 2006).

D. DATA COMPRESSION

The large number of sensor data is being generated in every several hour and needs to eliminate the redundancy by compressing sensor data, which becomes one of the most challenging tasks (Saket Sathe *et al.* 2006). Based on the accuracy in the sensor data is approximated, resulting in compressed representations of the data.
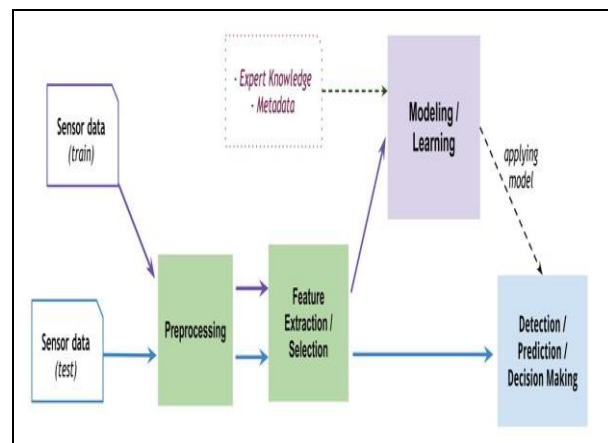

Fig. 2: Broad categories of sensor data management tasks

7. APPLICATION

The sensor data mining is used in various fields which is helpful to find the data and extract the useful knowledge (Hansen *et al.* 2006). Some of areas are social sensing applications and mobile data, software bug tracing in sensor networks, health care application, and environment and forecasting detection etc.

8. CONCLUSION

The sensor data model is based on the sensing data. A Sensor data Mining system (SDMS) is designed based on the model. This model service the useful knowledge according to continuous queries based on gathered information from sensor. In future we can implement the sensor data mining techniques which can be operated on this model.

FUNDING

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

CONFLICTS OF INTEREST

The authors declare that there is no conflict of interest.

COPYRIGHT

REFERENCES

Aggarwal, C. C. and Han, J., A Survey of RFID Data Processing, Managing and Mining Sensor Data, Springer (2013).

Aggarwal, C. C. and Yu, P. S., A survey of synopsis construction algorithms in data streams, Data Streams: Models and Algorithms, Springer (2007).

Aggarwal, C. C., Managing & Mining Uncertainn Data, Springer (2009).

Gaber, M. M., Zaslavsky, A. and Krishnaswamy, S., Mining Data Streams: A Review, In ACM SIGMOD, 34(2), 18-26 (2005).

Golab, L. and Ozsu, M. T., Issues in Data stream management, In SIGMOD Record, Vol. 32(2), 05-14 (2003).
https://doi.org/10.1145/776985.776986

Hansen, H. S., Loibl, W., Peters-Anders, J., Vogt, J., Zudin, S., Shuck, A., Varis, S. and Zaliwski, A SENSOR data management system overall design (2006).

Peng, Y., Kou, G., Shi, Y. and Chen, Z., A Systemic framework for the field of data mining and knowledge discovery, In Proceedings of Workshop on The 6th IEEE International Conference on Data Mining (2006).
https://doi.org/10.1109/ICDMW.2006.24

Saket Sathe, Thanasis G. Papaioannou, Hoyoung Jeung, Karl Aberer, A survey of model-based sensor data acquisition and management, 09-13 (2013).