

Research on Data Mining for the Internet of Things

Summarized by

Nur Amalina binti Razali

Faculty of Computing & Informatics

Multimedia University, Malaysia

Abstract

Implementation of data mining models in Internet of Things(IoT) is an efficiently method to store the massive data. Several methodologies are discussed based on the data mining models. This paper summarize four data mining models for IoT and how the models work in order to store and collect the data.

1 Problem Solved

The definition of Internet of Things (IoT) is hard to define in precise in term of concept itself. IoT is an open computing concept where it can describe the connectivity of physical objects, animals or people and be able to identify themselves as they have unique identifier. IoT bring a huge wave in world after revolution of computer and Internet because its ability to transfer and receive data over network is without requiring interaction between human and human and interaction between human and computer. The researchers make an implementation for IoT by implementing data mining models which are to reduce the risk of problem of data mining and IoT is lack of models to direct its data mining.

2 Claimed Contributions

The research introduces an approach of IoT using data mining models. The advantage of this approach is how the models implements on IoT although they focus on the foundation of IoT. Besides, this method can scale well with the numerous types of data in IoT.

3 Related Work

A brief overview of some related literature about data mining in the IoT, involving three important aspects of work. Firstly, works that concern on mining and managing the Radio Frequency Identification (RFID) stream data. . (?, ?) come out with RFID-Cuboid for warehousing RFID data. RFID data maintains three table which are info table, stay table and map table. These three table have their own function and characteristic. The Info table contains path-independent dimension for associated product while the stay table contains the TimeIn and TimeOut information for the different items that stay together at a location. The map table contains the hierarchy path information in order to do analysis of structure-aware for each item. The movement trail of RFID data form gigantic commodity flow graph and (?, ?) support Flow Graph to demonstrate the movement of commodity and analyze the flow of commodity.

Besides, there are works that concern the importance of querying ,analysing, and mining the moving object data. IoT able to generate many type of devices. (?, ?) introduced ROAM (Rule and Motif-based Anomaly Detection in Moving Objects)to detect the anomaly in moving object. (?, ?) introduces TraClass for trajectory data based on two types of clustering which are region-based and trajectory-based.

There are some reviews about discovery in sensor network based on the science and theory from

sensor data that explain the data mining in sensor network. (? , ?) introduces Spatio-Temporal Sensor Graphs (STSG) for modelling and mining the sensor. STSG is able to overcome the constraints in data as it can detect many types of motive including motive in anomaly.

4 Methodology

The method of the research is based on the four models which are multi-layer data mining model, distributed data mining model, grid based data mining model and data mining model for IoT from multi-technology integration perspective.

4.1 Multi-layer data mining model

This model contains 4 layers. The first layer is data collection layer, the second layer is data management layer followed by the third layer which is event processing layer. The last layer is data service layer. This modelling is based on the design of IoT and data mining of RFID.

4.1.1 Data collection layer

The layer adopts tools to collect the data of smart object. For example, RFID stream data can be collected from RFID readers. Every data requires a different data collection techniques. There will be a sequence of problems occur during the process of collection of data and the problems need to be well-solved.

4.1.2 Data management layer

The layer manages the collected data from data collection layer. The layer implements distribute database or data warehouse in order to collect the data. For example, one of the raw format of RFID data stream is EPC (Electronic Product Code). EPC is marked as ID of smart objects because it contains a unique identifier.

Along the process, Stay table provides the format records. RFID-CUBOID as a data warehousing is able to store and adopt the data. Apart from that, XML language provides data description. The users can online the process of data well.

4.1.3 Event processing layer

The function of event processing layer is to analyze programs in IoT. The program itself can provide a solid system to process the data. Based on that, the analysis can be done. As a result, data can be organized and filtered accordingly.

4.1.4 Data mining service layer

Event processing layer and data management layer are two important layers to form the data mining service layer. This layer is an architectural pattern in which the components provide the service to other components in order to make it as a service-oriented.

4.2 Distributed data mining model

Data in IoT is mass, distributed, time-related and position related. The sources of data is heterogeneous and limited of data resources. These features lead to some problems. First of all, difficulty for mining distributed data due to massive data is being stored in different place. Next, requirement for hardware of central node is high when adoption of central architecture is done because the data in IoT needs to be pre-processed. After that, the planning of putting the relevant data together is unworkable for some consideration towards data security and the other factors. Last but not least, the planning of sending the data minimize the use of energy and cost for transmission because of limited resources of nodes. The solution is pre-process the data in the distributed nodes and submit to receiver.

Distributed data mining model can break down the complex problem into simple ones, provide solution for the problems from distributed nodes as well as reduce the power of computing, high

storage capacity and performance. This model consists of grid middle ware, software resources, hardware resources, and components developed by data mining grid.

4.3 Grid based data mining model

The computation resources and power resources of Grid are beneficial for the user because the user be able to access abundance of resources which are computing resources, raw fact resources and tool resources. DataMiningGrid-based data mining model for IoT is different with DataMiningGrid in term of software and hardware resources part because IoT caters more software resources.

4.4 Data mining model from multi-technology integration perspective

The context-aware Intelligent Service Network provide the data to be collected. The connection via ubiquitous, for example RFID to access the Internet. A numerous of existing approach are provided to access next generation Internet. The authors finalize the data mining tools and algorithms. Then, based on this basic, the authors send the valuable knowledge to service-oriented application.

5 Lesson Learned/Experiments Details

Some considerations need to be taken into account for collection of data from smart objects of IoT. A strategies planning is important for data transmission and energy utilization of sensor nodes. There are limitations and constraints during the process. One of the characteristics of IoT is massive data. The management of data and implementation of online analytical query and processing must be well-managed and well-implemented. It is important to admit the semantics of data whenever the data is tend to have its own implicit semantics.

During the event filtering, it is important to monitor the process to ensure the data are aggregated based on the event. The business logic can be detected by detecting a complex event. The difference between centralized data processing and mining with distributed data processing and mining is minimization of the use of the transmission of energy and cost. The pre-process data for each distributed nodes is a better approach.

Algorithms of data mining include classification, clustering, outlier detection, and temporal patterns mining. Moreover, the next generation of Internet has many talent towards to development of technologies. Thus, it is important to study and recognize the problems that might be occur in data mining.

The experiment draws attention to a few problem points of this approach. Firstly, the models and theories for data mining is limited. Last but not least, a better study and investigation needs to carry out for the models suggested because of the characteristic of IoT itself.

6 Conclusions

This paper explained the data mining models to implement on the Internet of Things. Researching the current problem and preparing the methodology of this method are shown. First, the multi-layer data mining model which explains how the data is being collected and processed for each layer. Followed by distributed data mining model, grid based data mining model and lastly the data mining model for IoT from multi-technology integration perspective. This research paper contributes on introducing a new approach for each model on how the data works.

7 Future Works

The approaches used has a few limitations which are the characteristics of IoT makes the problems in data mining become a challenge task, limited of theories and models.

For potential future work, the generation should make a study about algorithm for IoT and test the algorithm by carrying out the experiment in order to provide a better model for IoT. Lastly, the Grid-based data mining

system and the corresponding algorithm can be tested out to get a new result. These potentials are a better approach to move forward and extend further in next generation of Internet.