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

Internet of Things(IoT) is a next technological transformation that can integrate new technologies of communication and communication. (?, ?) IoT is a integration of sensor network from technology perspective while it is an open concept where it can integrate new related technologies and applications from point of economy. The implementation of data mining for IoT is based on the data mining models which are to reduce the risk the problem of data mining but 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

Currently, there are three aspects of works about the data mining in the IoT.The first one is there are some works focus on managing and mining the Radio Frequency Identification (RFID) data stream. (?, ?) 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. Info table stores the path-independent information product while stay table stores the information about the item that stay together at a location. The function of map table is to store path information in order to perform structure-aware analysis. (?, ?) adopts Flow Graph to represent transportation of commodity and analyze the flow of commodity.

The second one is works that focus on interest in query,analyze and mine moving object data. IoT able to generate many type of devices. (?, ?) proposed ROAM for anomaly detection in moving object. (?, ?) put forward a novel partition and detect framework for trajectory outlier detection of moving object. He introduce TraClass, a new trajectory classification method. TraClass is a hierarchical region-based and trajectory-based clustering.

The other works are discovery of knowledge from sensor data. Based on the characteristic of sensor network, the data mining in sensor network has its own features.As a reference, (?, ?) introduce Spatio-Temporal Sensor Graphs (STSG) for modeling and mining the data.

4 Methodology

The method of the research on data mining models for IoT has four model 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

Multi-layer data mining model contains 4 layer : data collection layer,data management layer, event processing layer and data mining service layer. This modeling is based on the architecture of IoT and RFID"s data mining

4.1.1 Data collection layer

In data collection layer, it adopts devices to collect the data of smart object. 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

Data management layer is to manage the collected data from data collection layer. The layer implements distribute database or data warehouse in order to collect the data. For example, raw format of RFID data stream is EPC, location and time. EPC is marked as ID of smart objects.

Along the process, Stay table provides the format records. RFID-CUBOID as a data warehousing is able to save and manage 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 events in IoT. The event itself can provide a high-level mechanism for data processing of IoT. 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 layer for building the data mining service layer. The layer is built based on these two layers.The layer is service-oriented.

4.2 Distributed data mining model

Data in IoT is mass, distributed, time-related and position related. The sources of data of IoT is heterogenous 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 - processing of data in the distributed nodes and submit to the receiver.

Distributed data mining model can break down the complex problem into simple ones, solve the problems from distributed storage of nodes as well as reduce the power of computing, high storage capacity and performance.In this model, there are four important component which are grid middle ware, software resources, hardware resources and components developed by data mining grid.

4.3 Grid based data mining model

The user can make use of computation resources and power resources of Grid. Besides, The user be able to access abundance of resources which are computing resources, data resources and devices resources. There are differences between DataMiningGrid-based data mining model for IoT and DataMiningGrid. The differences are software and hardware resources part because IoT provides more software resources.

4.4 Data mining model from multi-technology integration perspective

In this model, the data are collected from the context-awareness of individuals, smart objects or environment. The model adopt 128-bit IPV6 address. A numerous of ubiquitous approach are provided to access next generation Internet.The authors finalize the data mining tools and algorithms.Then, 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 recognize 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 energy-costly transmission. In advance, pre-process data for each distributed nodes is a better approach.

Data mining algorithms for IoT include classification, clustering, outlier detection,and temporal patterns mining. Moreover, the next generation of Internet has many potential 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. First, IoT is lack of models and theories for data mining. Secondly, 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 intergration perspective. This research paper contributes on introducing a new approach for each model on how the data works.

7 Future Works

This approach has a few limitations: the characteristics of IoT makes the problems in data mining become a challenge task , limited of theories and models.

For potential future work, studying various data mining algorithm for IoT is a better work in order to provide a better model for IoT. Lastly, the implementation of Grid-based data mining system and the corresponding algorithm for able to move forward and extend further in next generation of Internet.