International Journal of Engineering Science Invention (IJESI)
ISSN (Online): 2319 – 6734, ISSN (Print): 2319 – 6726
www.ijesi.org || PP. 44-48

Difficulties and Data Mining Model For Internet of Things (IoT)

S.Maria Selvi[1]
[1](*Assistant Professor, Department of Computer Applications, St.Anns College of Arts and Science (Co-Ed), Tindivanam)*

Abstract: *The Internet of Things (IoT) is an emerging topic in today's era. It has a lot of significance in technology, business, social and engineering fields. This technology provides an easier way of communication of devices with the minimal interaction of human. IoT takes advantage of advancement in network interconnections and computing ability to propose new techniques. Data mining techniques are used to handle the massive data generated by the Internet of Things (IoT). Various data mining models have been proposed for Internet of Things. We are presenting a novel data mining model for Internet of Thing which considers typical IoT challenges. In this paper we are focusing on the research issues of IoT with respect to the data mining and various data mining models for the Internet of Things (IoT).*

Keywords: *Internet of Things, Data mining models.*

I. Introduction

The Internet of Things (IoT) refers to the next generation of Internet which will contain trillions of nodes representing various objects from small ubiquitous sensor devices and handhelds to large web servers and supercomputer clusters. The large scale implementation of Internet of Things devices promises to transform aspects of the way we live. For consumers, new Internet of Things products like Internet enabled appliances, home automation components, and energy management devices are moving us a towards a vision of the "smart home", offering more security and energy efficiency.

Other personal Internet of Things devices like wearable fitness, health monitoring devices and network enabled medical devices are transforming the way health care services are delivered. This technology promises to be beneficial for people with disabilities and the elderly, enabling improved levels of independence and quality of life at a reasonable cost. According to the contractual agreement, Internet of Things connects any items with Internet, implements information exchange and communicates through information sensing equipment such as the sensor, radio frequency identification (RFID) and global positioning system. Internet of Things needs be designed to identify, supervise and monitor items so as that it can provide various type of information services for users innovatively. The Internet of Things has very complex data types, including sensor data, radio frequency identification data, two dimensional code, video data and image data. The data in Internet of Things can be categorized into several types: radio frequency identification data stream, address/unique identifiers, descriptive data, positional data, environmental data and sensor network data etc. The brings the great challenges for managing and analyzing and mining data in the Internet of Things.

Huge amount of data are generated by environmental sensors every day across the globe. There is tremendous need for data analysis system which are able to mine massive and continuous stream of real world data applications such as temperature monitoring, air pollution, stock market and network security etc. Data generated by environmental sensors are recorded at time intervals of second through to minutes and over time these sensor will create dataset that need to be mined in real time in a way that takes into create consideration the dynamic natures of the real world changes that are being measured.

Data mining is the process of posing quarries and extracting patterns, often previously unknown form large quantities of data using pattern matching or other reasoning techniques. This paper mainly focuses on the data mining challenges and data mining models for Internet of Things (IoT).

II. Related Works

Internet of Things (IoT) with data mining currently becomes the major research area of researchers. Various data mining concepts have been proposed for IoT. However, data management and technical issues and challenges have emerged in recent years. In this paper we have focused on these challenges and data mining models for Internet of Things (IoT).

Data mining involves knowledge discovery and analysis from the massive set of data. The main purpose of data mining is to find useful patterns from large dataset received from Internet of Things (IoT) devices, sensors. Knowledge discovery, pattern analysis and information harvesting are the terms which are used for data mining in Internet of Things. The primary objective of data mining is to build an efficient and descriptive model which is best suitable for the data set.

In the recent years various data mining techniques and models have been published for Internet of Things. Further we will discuss the data mining models and challenges in IoT.

A. Data Mining Models:

Data mining is the process of extracting useful information or patterns from the raw data. Data mining in internet of Things is used to manage the large amount of data which is received from the IoT devices. On the basis of definition of data mining and its functions, a typical data mining process includes these four steps.

1. Data Collection
2. Data preparation
3. Data mining
4. presentation

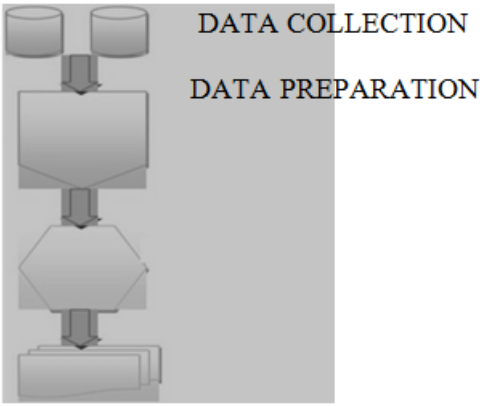

Fig.2 IoT with Knowledge Discovery Model

Knowledge Discovery Model which shows how the raw data has been converted into useful information. The second data mining model is the multilayer model in which each layer of the model does their respective tasks like data collection, data processing and data mining services. The third data mining model is for distributed architecture which serves and analyses the data stored in the distributed locations. The fourth data mining model is grid based model which utilizes the high amount of data or we may say unlimited data optimally.

1. IoT with Knowledge Discovery Model

The IoT devices generate a massive amount of data. The Knowledge Discovery Model (KDD), when applied to IoT, Converts data collected by the IoT devices into useful information which is used to generate knowledge. This model is very similar to the data mining knowledge discovery model. It consists of following steps. First is IoT database, which consists of sensor data, application data and data from other

The data mining step is solely responsible for extracting useful patterns from the data obtained after the data processing step. The result of data mining process is then used in the decision making step. The decision making step converts the mined data into a useful set or knowledge. All the steps of KDD process is required to obtain the efficient data. It is also important to note that the data fusion, large scale data, data transmission, and decentralized computing issues may have a stronger impact on the system performance and service quality of IoT than KDD or data mining algorithms alone may have on the traditional applications. It is also important that large scale data, data fusion, data transmission and distributed architecture computing issues may have a strong impact on the system performance and service quality.

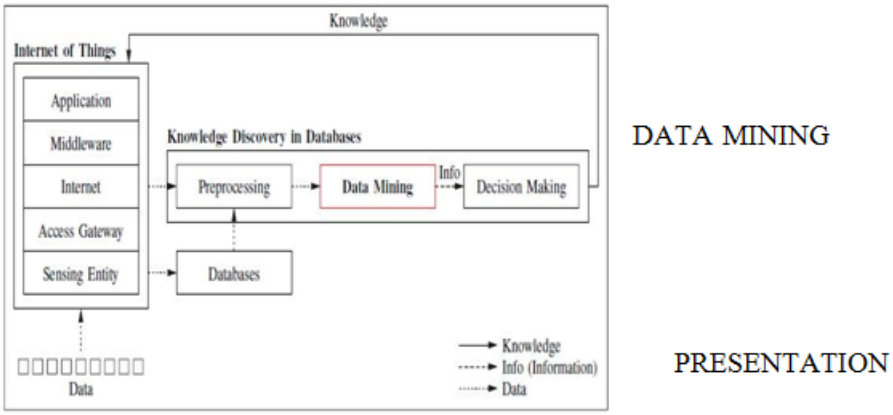

Fig.3 Multilayer Data Mining Model for IoT

2. Multilayer Data Mining Model

The multilayer data mining model is divided into six parts. The Data collection layer consists of sensor data and quality data. Next layer is data management layer which manages data from distributed, central and data warehouse. Next layer is event process layer which provides the integration of data management and control. It provides the high enquiry based data. The event filtering is the most important step in this layer. Another layer is event mining service layer which provides the facility of data cleaning and merging. In this layer data mining operations are selected to obtain the efficient solution to the problem. The mined data is used to obtain knowledge.

This model consists of two main controls i.e.; vertical control and horizontal coordination. These layers are used to provide management and control to the mining process on local nodes and then it is made visible to all the nodes by updating the global node.

3. Data mining model for distributed architecture

The above two models are mainly based on handling local data. This model is mainly described for the distributed infrastructure. This data mining model deals with the data coming from the different nodes.It consists of two main layers global node and local node. A distributed architecture consists of multiple nodes in different sites. Each node in the architecture consists of a local node and a global node. The management of data must be in the local site as well as in the global site. To manage and mine data in distributed environment this model helps to mine data in both the sites. Data mining operations are firstly performed on local nodes and then it is made visible to all the nodes by updating the global node.

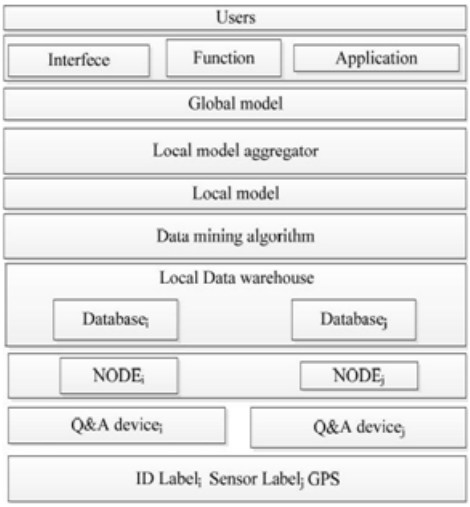

Fig.4 Data mining model for distributed architecture

4. Grid based data mining model

Grid computing provides computational capability in heterogeneous environment. The Internet of Things environment provides collaboration of different environments parallel. The grid based data model provides the facility of parallel task execution or parallel data mining.

The model is also known as grid miner which is used to mine multi model sensor data and used to combine them. It is mainly consists of five layers namely IoT resource layer, IoT Service layer, grid middleware layer, grid mining layer and grid application layer. The major difference between Data mining model in IoT and Grid Data mining can be a part of hardware and software. It provides various.

Hardware supports such as RFID, WSAN and WSN as well as sensor nodes.

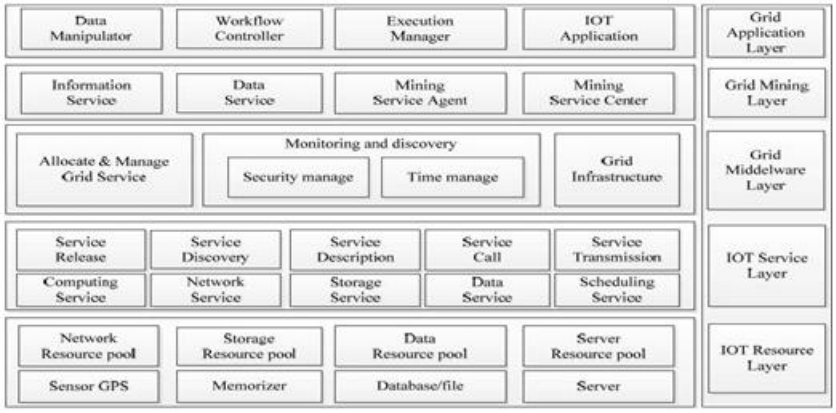

Fig.5 Grid based Data mining Model

III. Difficulties Of Data Mining For IOT

We have reviewed following challenges in Internet of Things with respect to the data mining.

1. **Security**: As the Internet of Things connects more devices together, it provides more decentralized entry points for malware. More layers of software, integration middleware and machine to machine communication etc. create more complexity and new security risks. Expect to see many different techniques and vendors addressing these issues with policy driven approaches to security and provisioning.
2. **Data Mining Algorithm:** This is the most important issue as it is necessary to select a suitable data mining algorithm.
3. **Selection of Data Mining Model**: An IoT device generates a large amount of data. To handle and manage this data we need to select an appropriate data mining model for it.
4. **Accessing and Data Extraction:** Data extraction from the large data set is a primary challenge of IoT. A proper mechanism is needed to extract data from the massive data set.

Handling of Heterogeneous Data: As IoT is all about heterogeneity. Data must be collaborated from different environments. Handling of data coming from distributed

IV. Proposed Data Mining Model

After considering all the aspects of data mining and its application and challenges in Internet of Things we have proposed a data mining model. The model consists of five layers.

1. Data collection layer
2. Data management layer
3. Data analysis layer
4. Data processing layer
5. Data mining
6. Data mining layer provides all the functionalities.

V. Conclusion

In our proposed model we have defined following functionalities of each layer.

Data Collection layer collects data from sensor data, camera, RFID and other storage whether it is from local site and global site.

Data Management layer manages data coming from local site or from a distributed site. This layer deals with IoT data coming from different environments.

Data Analysis layer deals with the type and structure of the data. It analyses the IoT data whether it is real time, current or historical data. It is very important to manage these types of data separately to obtain efficient output.

This layer deals with the structuring of the data as we may get data in any form as structured, semi structured or unstructured. The important task of this layer is data parsing. It parses data based on system and requirement of the data.

Data Processing layer deals with the preprocessing of data which we are getting after data analysis. This layer includes three steps. First is **Data Cleaning** in which data is checked and processed for noise removal and unknown values.
Second is **Dimensionality Reduction** which reduces dimension by using machine learning techniques Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA).

Third is **Merging,** in which different IoT data are merged together. This step is important because Internet of Things is all about collaboration of different data coming from IoT devices.

Data Mining layer is the fifth and last layer of our data mining model for the Internet of Things (IoT). This layer performs all the necessary data mining operations on data. A suitable data mining algorithm is selected for the data getting after the preprocessing step. This layer helps to generate useful result or knowledge from the mined data.

So, the model describes all the steps of mining massive data coming from the Internet of Things (IoT) environment. This model also deals with the networking of the data and check data is coming from which environment. Data can be accessed from a local database, centralized server and distributed database. All the five layers of this model have its own significance. To obtain an optimal result raw data should pass from all these layers in a sequence.

PROS:
1. Consists all the primary steps.
2. Considers data coming from multiple environments.
3. Analyses data whether it is structured, semi structured and unstructured.
4. Data Parsing is its feature which is useful in distributed environment.
5. This model provides all the features of data preprocessing.

References

[1]. Tsai, Chun-Wei, et al. "Data mining for internet of things: A survey."Communications Surveys & Tutorials, IEEE 16.1 (2014): 77-97.
[2]. Stankovic, John. "Research directions for the internet of things." Internet of Things Journal, IEEE 1.1 (2014): 3-9.
[3]. Bin, Shen, Liu Yuan, and Wang Xiaoyi. "Research on data mining models for the internet of things." Image Analysis and Signal Processing (IASP), 2010 International Conference on. IEEE, 2010.
[4]. Zhang, Chunguang, et al. "Analysis on Data Mining Model Objected to Internet of Things." International Journal of Advancements in Computing Technology 4.21 (2012).
[5]. Han, Jiawei, Micheline Kamber, and Jian Pei. Data mining: concepts and techniques: concepts and techniques. Elsevier, 2011.
[6]. Gonzalez, Hector, Jiawei Han, and Xiaolei Li. "Mining compressed commodity workflows from massive RFID data sets." Proceedings of the 15th ACM international conference on Information and knowledge management. ACM, 2006.
[7]. Gao, Kun, Qin Wang, and Lifeng Xi. "Controlling Moving Object in the Internet of Things." International Journal of Advancements in Computing Technology 4.5 (2012).
[8]. Yan-mei, Ma, and Ren Hong-e. "Application of RFID and Data Mining in the Timber Management System." Control,AutomationandSystems ngineering (CASE), 2011 International Conference on. IEEE, 2011.
[9]. Gupta, Misha, and Er Jyoti Arora. "A review on major problems in Internet of things implementation."
[10]. Gubbi, Jayavardhana, et al. "Internet of Things (IoT): A vision, architectural elements, and future directions." FutureGeneration Computer Systems 29.7 (2013): 1645-1660.
[11]. Saini, Dinesh Kumar, and Sanad Al Maskari. "Data Management Issues and Data Mining of Real Time System Application for Environment Monitoring." (2014).
[12]. Aggarwal, Charu C., Naveen Ashish, and Amit P. Sheth. "The Internet of Things: A Survey from the Data-Centric Perspective." (2013): 383-428.