A Study on Use of Data Mining Techniques in IoT

Annu Mishra

BBA Department, JIMS,IP University

***Abstract*— Internet of Things (IoT) has grown drastically due to research progressions and various technological advancements. Communicating and linking objects together through internet seems to be very difficult, but Internet of Things (IoT) has made it possible for us. In today's modern technical world the IoT is playing a vital role in connecting different domains together for better analysis and vision of information. As IoT collects a large volume of data, it is very much required to extract new data from the old one and thus perform data mining for useful information from the ocean of data. The data mining models used along with the Internet of Things can really lead to new conclusions. In this paper, I have presented various data modelling models and discussed about its usage in IoT. I have also mentioned its advantages and disadvantages. In the end I have also mentioned the challenges that IoT faces.**

***Keywords*— Internet of Things, Data Mining Models, Knowledge Discovery in Databases, Application of Data Mining**

I. INTRODUCTION

The Internet of Things (IoT) can be defined as the network which is capable of connecting anything like vehicles, buildings, objects and equipment having network protocol support system, sensors, software and advanced technology. It is thus sometimes referred as IoE i.e Internet of Everything. These features not only help in collecting data but also helps in controlling and analysing the data. The network and the IoT together boosts the collection and controlling huge amount of data. So to make IoT more advanced and vital, data analysis is required and the best way for data analysis is Data Mining Data mining refers to the process of analysing databases for gaining new information out of the old data stored. This type of mining lays the information which is new, dynamic, useful and advanced. It is also known as Knowledge Discovery in Databases (KDD). Knowledge Discovery in Databases includes problem identification, data elicitation, data pre-processing, data transformation, selecting data mining model and final result evaluation or analysis. Thus KDD is a repetitive cycle. Data mining can be used with several domains like algorithms, machine learning, artificial intelligence, data warehouses, databases, data handling etc. One of the limitations of IoT is that the sensors involved in the IoT are very complicated and it gives large amount of heterogeneous data which needs to be cleaned and filtered so that it can be presented to the user in a clear and appropriate manner and thus the user will be able to have an insight on the information provided. Assessments recommend that IoT will be an accumulation involving roughly 50 billion objects before the finish of 2020 [4]. Data mining can therefore be defined as a process of investigating large and voluminous data to find new and valuable information. Data mining can thus prove to be a revolutionary domain in extracting new information and then use it for all other social and technical aspects.

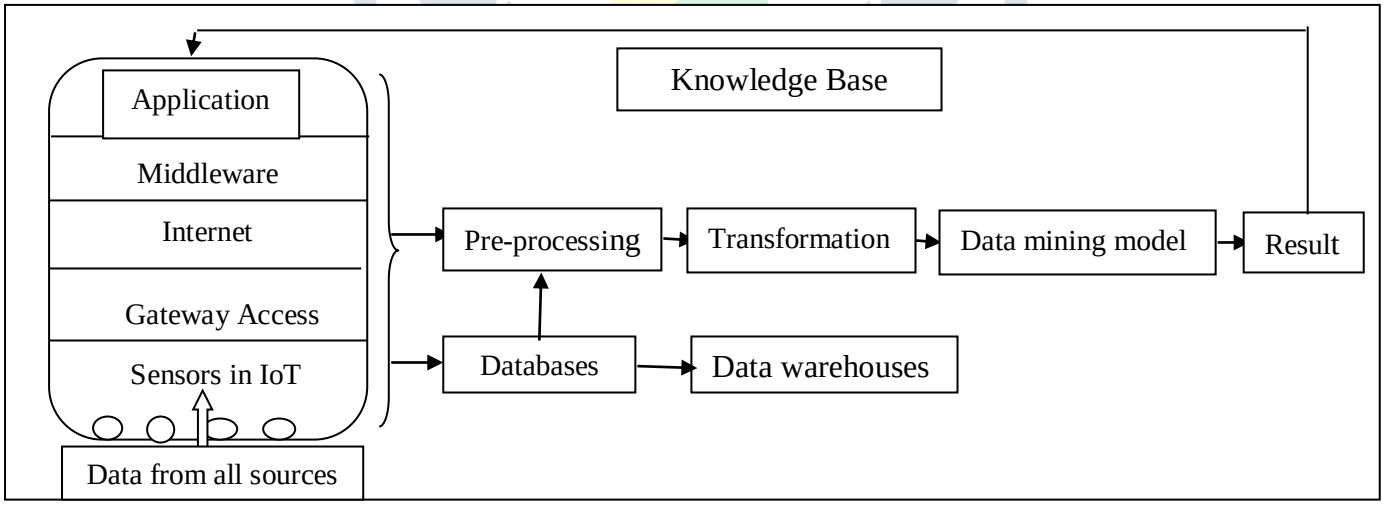

Figure 1. Collection of data by IoT

The Figure 1 above clearly shows how IoT gathers information from various sources, which may or may not contain information for other systems and IoT itself. Now we know that IoT collects huge amount of data and thus if the KDD or Data mining is applied to this vast amount of data it will definitely lead to some useful or valuable information which could be used for further learning process. The figure shows step by step process of collecting data, pre-processing, transformation and result analysis. This result gained can be used again by the IoT to as a valuable information and when combined with the KDD it may lead to much more usable information. However it must be noticed that not all the data collected by the IoT is useful for data mining as the process of data mining also depends upon the placing and timing of the data. In this paper I have included basic architecture of data mining in collaboration with the IoT, several data mining models used in IoT and use of data mining in IoT. I have also mentioned the works done together with IoT and data mining and limitations of the data mining as well.

II. RELATED WORKS

Not much has been done in the field of IoT as of now as it is a new domain of research. However, it is trending to be used in domains like navigation systems, radio frequency identification, agriculture for automating the irrigation, healthcare, defence, manufacturing etc. It can discover inconsistent patterns, incorporated areas at each time interim, and even hubs eligible to be future hotspot. Smart homes are built with the help of IoT and using mechanisms to control the internal temperature of the house according to the outside ambience. It is also used in the retail sector where the path of the customer can be tracked and the store can be filled with premium products in heavy traffic arears. In [13], the author has proposed a design for high-performance data mining module suitable for IoT with some basic prerequisites. These requisites are selection of objective, data characteristics, and the mining algorithm suitable for it. For selecting the objective, it is required to select the relevant mining technique for the given problem domain as there are several mining techniques. The second prerequisite is Data characterization for which the size of data and its representation matters a lot. Data can be Automation data, Status data or Location data. Automation data is the type of data related to the automated concept like switch off the lights when not at home or adjust the geyser temperature remotely. Status data are those data which depicts the status of the object and location data refers to those data which are generally based on the positioning data. Different types of data must be processed in distinct way. It may happen that data collected from various sources are same in some form but it must be processed distinctly based on the system from which it has arrived as these data have different implications on the system. The third and last pre-requisite is the Mining algorithms to be used. It is somewhat easy to choose the right mining algorithm if the objective and the data representation is clear. Some authors prefer developing new algorithms while some may use the existing one.

III. TYPES OF DATA MINING MODEL USED FOR IOT

A. *Multi-Layer Data Mining Model*

In Multi-layer data mining model, the model is divided into four layers which are data mining service layer, processing layer data management and data collection layer. In the data collection layer some devices and equipment like radio frequency identification sensors, GPS sensors, satellite positional and many other devices and equipment embedded with IoT. Some of this information may be highly sensitive and complex too. These types of data require special storage methods. Data collection is one of the basic and most important steps in the area if data mining as it tends to provide voluminous data for the data mining and to gain new information out of the stored data. There may be several problems during this step like data pre-processing, communication and format of the data. The second layer is the Data management layer where the data collected in the previous layer is stored properly in the databases and data warehouses. It is the place where the huge amount of data is stored for further analysis. Next layer is the Event management layer where the information from the databases or data warehouses are combined based on several key combinations which in turn creates a series of events. In the uppermost layer i.e. the Data mining service layer, the events are categorized and information is gained using complex queries. Each layer in the Multi-layer data mining model has its role and responsibilities. In the data management layer, we have the ID Tags which are very important for the identification purpose of the device or the object attached. Identifiers are the pattern which distinctly and uniquely identifies the entity. The ID tags can be categorised into three types namely Passive tags, Semi-passive tags and Active tags. The passive tags have less computational capacity and cannot detect the collisions as well. Passive tags cannot communicate among themselves. They can only react on the reader's command. Semi-passive tags possess the capacity recharge the microchips used in the devices. Active tags on the other hand can sense the channel and detect the collision. Another type o9f tag is the Sensor tag. These are the sensors which has the capability to convert any type of pressure into the electric signal. These sensor tags can be used along with the IoT devices. I have also shown the GPS tags which are based on the positioning data of the devices.

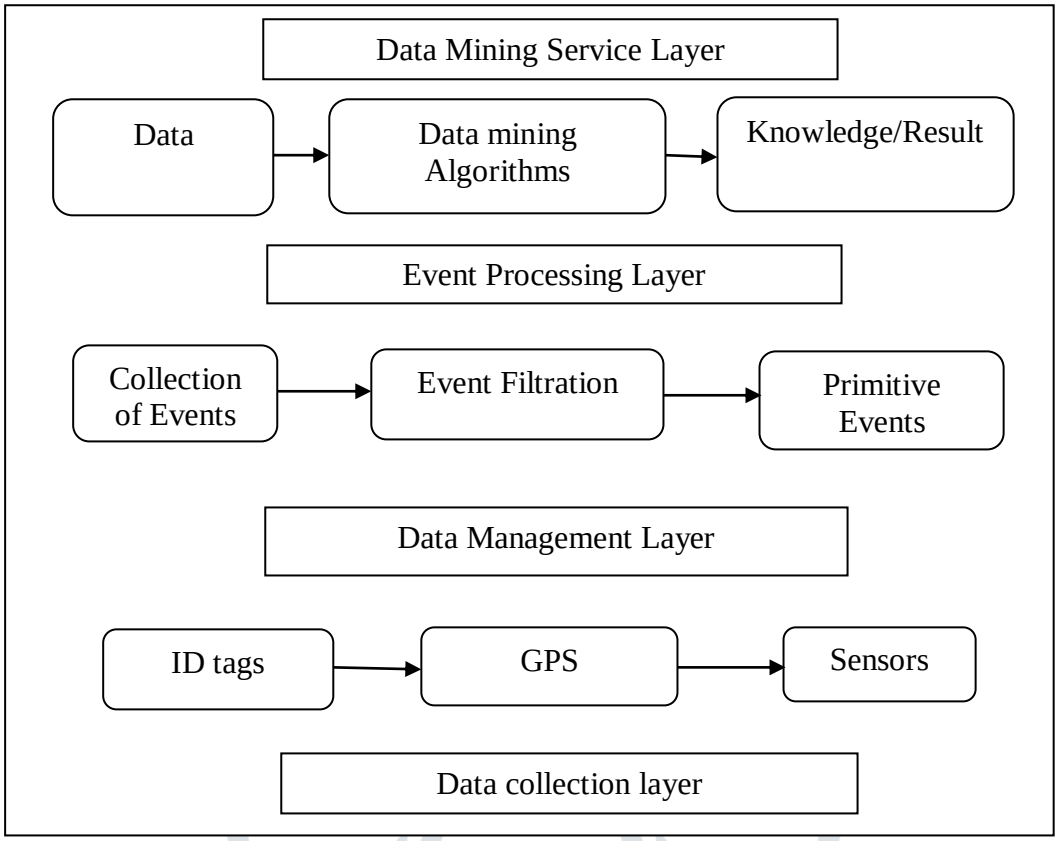

Figure 2. Multi-layer Data Mining Model

B. *Distributed data mining model*

The information in IoT has many attributes. Some of the information may be place and time dependent while some of them might have other relative attributes. Thus, in distributed data mining model it becomes very difficult accumulate the alike data for any type of analysis as the information may be heterogeneous in nature. In distributed data mining model, many multi-layer data mining models at different places may produce the result and these results are then accumulated at a common place namely the Global Model. Thus, the pre-processing of data is required at every stage of mining. Moreover, the main issue with the distributed data mining model is security of the data during the data transfer to the global accumulator. In Figure 3, I have described the model of distributed data mining model which clearly states how the data from different multi-layer data mining model are collected and analysed further or this information may also be useful for some applications as well. It can be noted that these multi-layer data mining models may also have distant format and procedures from each other thus the task of distributed data mining model increases. All the activities like data collection, data management, event management and data mining occur in this type of model too. The only difference between multi-layer and distributed data is that the data mining in distributed data mining model occurs at different places whereas in multi-layer data mining model the data mining process occurs at a single place.

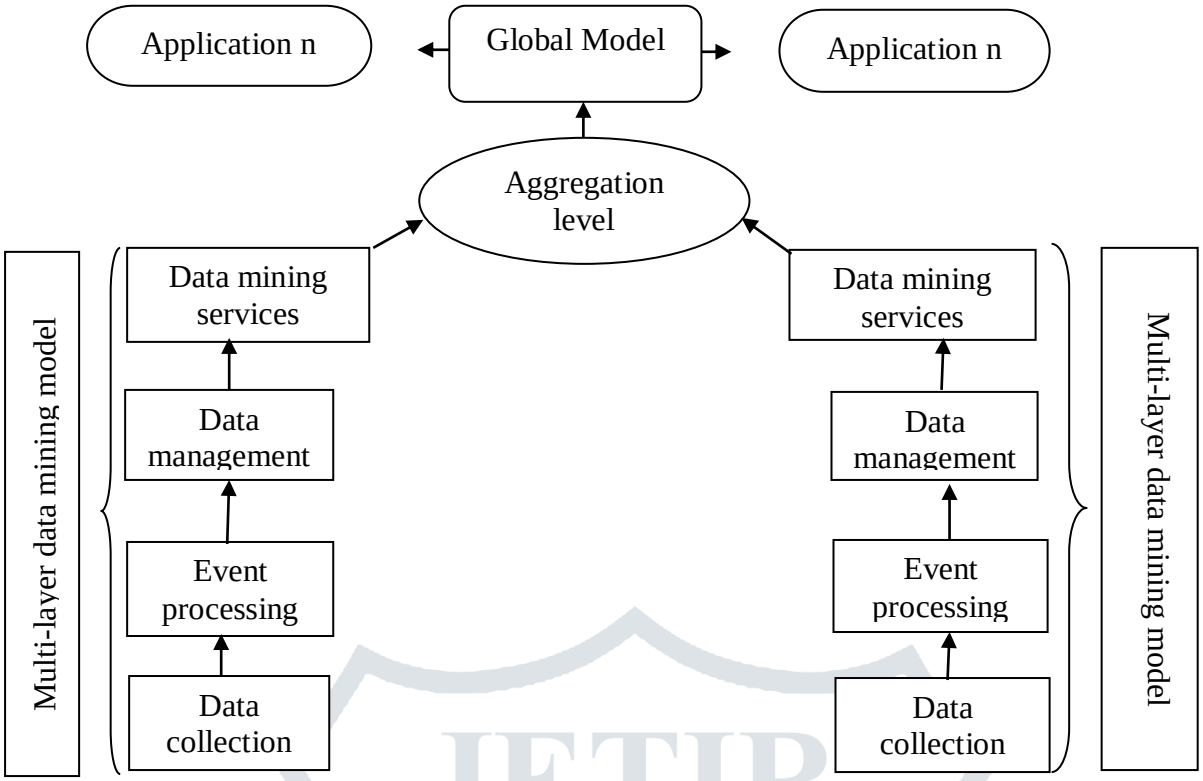

Figure 3. Distributed data mining model

IV. APPLICATIONS OF IOT

IoT is used in various domain and applications. Although there is much more to be explored about it, but in the recent years it has proved to be a revolutionary. Some of the areas where the IoT is applied are emotion modelling. There are some wearable gadgets which records the heart rate, pulse rate, face expressions and thus determines the emotion of the person over the period of time. Some of the giant retail sectors are making use of IoT to determine the recent trends and demand of the consumer and thus boost the production and supply of the product accordingly. The smart appliances and gadgets are making use of IoT to an extent. Some of the gadgets like watches, mobile phones, camera are carrying this application by default. This devices and gadgets record the data in the form of steps moved, calories burned, water intake etc. IoT is also playing a vital role in the field of agriculture. The farmers can predict the weather and crop growth in advanced. They can also get the advice for irrigation and cropping. Another area where the IoT is used is the Cloud computing. The data can be gathered in the cloud server and then the huge amount of data can be used to analyse the data and create new information.

V. ISSUES OF DATA MINING IN IOT

There are number of issues involved in data mining in IoT:

A. *Effective data elicitation*
As the data collection involved with IoT is vast and heterogenous, thus it is a challenging option to gather the data in proper format and fault tolerance must be considered.

B. *Aggregation of data*
The huge data collected by the IoT must be aggregated in a proper storage and proper format.

C. *Abstraction of data*
As the data collected is massive, some of the data may not be of use. Thus, data must be abstracted.

D. *Distributed data processing and mining*
As mentioned earlier, the data is heterogenous through IoT and thus it requires pre-processing too. Moreover, as it arrives from different sources, the mining process is also considered to be very challenging.

VI. CONCLUSION

IoT is going to play a major role in almost every industry in the near future. The versatility given by the IoT is going to be a boon for health, agriculture, engineering, defence and many more domains. The issues related with the IoT with and without data mining can be resolved with further research works and practices. The key features like spatial data, RFID and many more not only help in collecting data but also helps in controlling and analysing the data to gain new information and benefits. Thus IoT is highly recommended for the growth and improvement of the society.

REFERENCES

[1] Bijalwan V, Kumar V, Kumari P et al (2014) KNN based machine learning approach for text and document mining. Int J Database Theory Appl 7(1):61–70

[2] Olaiya Folorunsho and Adesesan Adeyemo. Application of data mining techniques in weather prediction and climate change studies. 4, 02 2012.

[3] Jeu Young Kim, Hark-Jin Lee, Ji-Yeon Son, and Jun-Hee Park. Smart home web of objects-based iot management model and methods for home data mining. In Network Operations and Management Symposium (APNOMS), 2015 17th Asia-Pacific, pages 327–331. IEEE, 2015.

[4] Saral Nigam, Shikha Asthana, and Punit Gupta. Iot based intelligent billboard using data mining. In Innovation and Challenges in Cyber Security (ICICCS-INBUSH), 2016 International Conference on, pages 107–110. IEEE, 2016.

[5] Dhakar M, Tiwari A (2014) A novel data mining based hybrid intrusion detection framework. J Inf Comput Sci 9(1):037–048

[6] Alexander Muriuki Njeru, Mwana Said Omar, Sun Yi, Samiullah Paracha, and Muhammad Wannous. Using iot technology to improve online education through data mining. In Applied System Innovation (ICASI), 2017 International Conference on, pages 515–518. IEEE, 2017.

[7] Sebastian Scholze Claudio Cenedese Oliviu Matei, Carmen Anton. Multi-layered data mining architecture in the context of internet of things. In IEEE. IEEE, 2017.

[8] Brijesh Iyer Prachi Deshpande. Research directions in the internet of every things(ioet). In International Conference on Computing, Communication and Automation (ICCCA2017), pages 1353–1357. IEEE, 2017.

[9] Okazaki S, Díaz-Martín AM, Rozano M et al (2015) Using twitter to engage with customers: a data mining approach. Internet Res 25(3):1066–2243

[10] Vangipuram Radhakrishna, Puligadda Veereswara Kumar, Vinjamuri Janaki, and Shadi Aljawarneh. A computationally efficient approach for temporal pattern mining in iot. In Engineering & MIS (ICEMIS), International Conference on, pages 1–4. IEEE, 2016.

[11] Xin Song, Cuirong Wang, and Jing Gao. An integrated framework for analysis and mining of the massive sensor data using feature preserving strategy on cloud computing. In Computational Intelligence and Design (ISCID), 2014 Seventh International Symposium on, volume 2, pages 337–340. IEEE, 2014.

[12] Dr. Shalini Bhaskar Bajaj Tushar Taneja, Aman Jatain. Predictive analytics on iot. In IEEE, pages 1312–1317. IEEE, 2009.

[13] Hai xiang Zhao and Fred´ eric´ Magoules.` A review on the prediction of building energy consumption. Renewable and Sustainable Energy Reviews, 16(6):3586 – 3592, 2012.