International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-4, November 2019

Techniques, Models and Challenges of Data Mining in Internet of Things (IOT)

M Chandra Prabha, R Viswanathan

Abstract— Due to many achievements in the field of communication and sensor networks, Internet of Things becomes an emerging technology which makes the human life easier and comfortable. This technology development paves the way to work and live in a comfortable manner. In internet, interfacing of any object is very difficult, but IoT makes the process very easier that too in a short span of time. In IoT, a huge of amount of data has been captured that is considered as huge business and social value data's'. So, IoT needs a model which is used to extract those high business data. Thus, data mining is a process which extracts high business data from huge amount of data easily. In this paper, a systematic review has been carried out on various data mining techniques and models that have been applied in IoT. Also its advantages and disadvantages have been discussed along with the challenges and future trends of IoT..

Keywords: Internet of Things, Data Mining, Data Mining techniques, Data Mining Models, Challenges.

I. INTRODUCTION

The tremendous growth of electronic devices such as smart phones, tablets that can be communicated either physically or wirelessly is a necessary activity in day to day life. This activity leads to the development of internet world is called Internet of Things which helps to connect the sensors, appliances, vehicles and other things easily and remotely. IoT helps the human to connect any of their devices from anywhere and at anytime. Also it provides the access for any objects efficiently. The main vision of IoT is to expand the benefits such as efficient data sharing, ability to control the remote devices, constant connectivity etc. In the field of communication and sensor technology, IoT connects an everyday objects through a global network.

Data Mining is known as a process of knowledge discovery from large databases because it analyzes the large datasets for pattern mining that are very novel, justifiable and easy to understand. Knowledge Discovery is an iterative process which includes the formulation of problem, collection of data, cleaning of data, and displaying the discovered data to the users. Also, Data mining has a link with other fields such as statistics, machine learning, AI, databases etc but the main goal of data mining is on the automation and handling of voluminous and heterogeneous data.

Due to the continuous integration of classical networks, IoT achieves efficient monitoring and controlling of all things which results into a generation or capture of large amount of data. In IoT, data's can be categorized into different types such as RFID data stream, unique address identifiers, descriptive data, positional data, environmental data, and sensor network data. These data's has been received from various sources like temperature monitoring system, air pollution control system, stock market, network security etc. These categories of data lead to a major challenges in managing, analyzing and mining of data's in IoT framework. Thus, data analysis such as clustering, classification and prediction is required to make the IoT smarter and effective. In recent years, most of the research has been focused on the application of data mining models in IoT. In this paper, various data mining challenges in IoT is reviewed. Also, various data mining models for IoT is studied along with its merits and demerits.

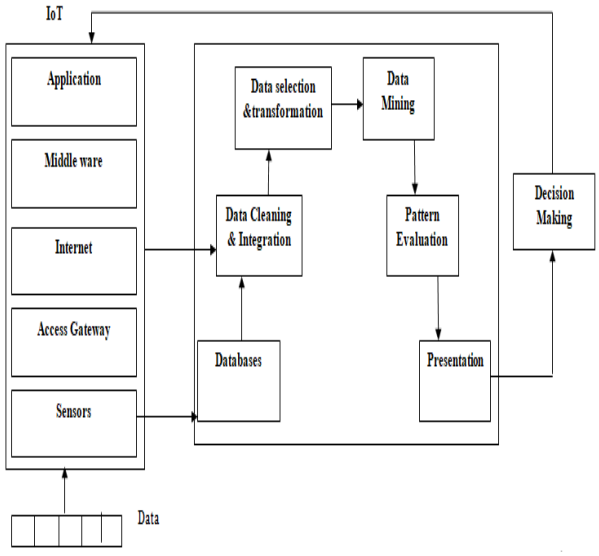

Fig.1. IoT with Data Mining Architecture

II. DATA MINING TECHNIQUES IN IOT

Data Mining becomes very important and trending technology in IT industry. This technology is mainly used to extract the certain needed data's and patterns from the heterogeneous data sets and that will be useful for the growing organization to improve their performance level because of the efficient data handling.

Manuscript published on November 30, 2019.
* Correspondence Author
M ChandraPrabha*, , Research Scholar, Galgotias University, Greater Noida, U.P, India.
Dr R Viswanathan, Associate Professor, Galgotias University, Greater Noida, U.P, India.

In databases, the main aim of the knowledge discovery is to find the innovative and
frequent patterns from the enormous amount of data. The combination of IoT and Data mining leads to the development of groundbreaking technology that makes a tremendous growth changes in IT industry. In IoT, data's can be differentiated into so many categories like "data about objects" and "data generated by objects". Data about objects is defined as the data that can be described by the objects itself. Whereas, Data generated by objects can be referred as the data that has been generated or captured by embedded devices. These data's are human interacted or human – system interacted data. There are various data mining techniques like clustering, classification, association etc. In this paper, some of the techniques are described below in view of IoT.

2.1 Clustering of data's in IoT

Clustering is a concept which groups the set of data's into clusters based on their similarity. These clusters are very helpful to analysis the data collected by the embedded devices of IoT.

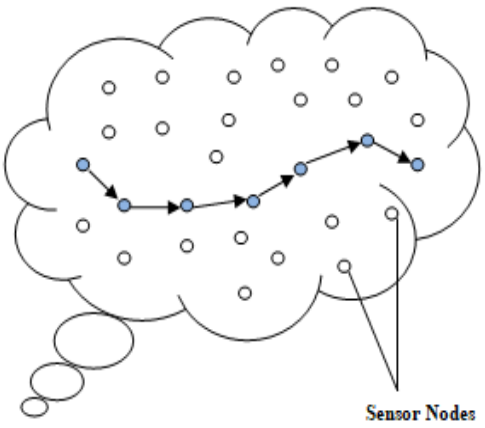

Fig.2. Clustered Nodes in WSN

Clustering and its algorithms are mainly used to overcome the challenges in managing the IoT data's and its objects. In IoT, WSN data's are highly energy conserving data's. LEACH (Low Energy Adaptive Clustering Hierarchy) is a hierarchical protocol that can be integrated with clustering of nodes in WSN. The main motive of LEACH is to reduce the energy consumption to create and manage the clusters of nodes and cluster heads in WSN. Also, IoT objects are grouped into different clusters that will help to overcome the problems like energy utility & its efficiency, scalability and robustness issues. Below figure depicts the clustering of nodes in WSN to reduce the energy conserved by sensor nodes. Following are the some of the clustering techniques used in data mining to cluster the data.

a) Partitioning clustering

This clustering technique helps to make cluster by the method of point relocation iteratively between several subsets or in heavily populated areas. K-means, k-medoids are the some algorithms for this clustering technique.

b) Hierarchical clustering

Hierarchical clustering merges data objects into subgroups. These subgroups are again merged into another level of larger sub groups and so on. There are two types of hierarchical clustering, namely

1. Agglomerative approach follows bottom up model.
2. Divisive approach follows top down model.

c) Density based clustering

It locates or identifies the clusters from the high density regions. The algorithms of this clustering are DBSCAN, GDBSCAN, and OPTICS.

d) *Grid based clustering*

It mainly focuses on space which has been bounded by more and more data points. IN the area of dense grids, grid based data structure is used to make clusters. The algorithms in this clustering are STING, wave cluster and CLIQUE

2.2 Classification of data in IoT

Classification plays a vital role in decision making process of collected data. Classification is a process of predicting and assigning an unknown and unlabelled data to predefined labeled dataset based on its features. For example, in credit card division of bank, identifying and classifying the set of customers as low, medium or high risky category, according to their salary and family background details. Classification has different methods. Some of the important methods are:

i) Decision tree induction

Decision tree can be used to classify the sensors used in IoT. Sensors are used to sense the different kind of data in various applications of IoT. Decision tree helps to predict the type of sensor that has to be used for that particular application like traffic jam prediction application, agricultural application, health care application etc.

ii) KNN Classification

KNN classification is called as K-nearest Neighbor classification method. It is used to find out the nearest nodes in WSN and helps to clusters it and reduces the energy conserved by the sensors used in WSN.

iii) Naive Bayes Classifiers

It is represented as directed acyclic graph. A node of the graph represents the random variables and edges represent the conditional dependency of the nodes. If nodes aren't connected to the graph, those nodes are called as variables but they are conditionally independent of each other.

2.3 Association of IoT Data

Association rules are defined as the if-then statements and it is used to analyze and predict the customer behavior. Otherwise, we can say that it shows the relationship between the item sets in large datasets from the different types of databases. In IoT, Association rules can be used in the field of agriculture. Example, from the agriculture dataset, we can able to analyze and predict the number of crops to be cultivated according to the monsoon and water availability. This can be done on the basis of previous success rate of crops cultivated during that period.

III DATA MINING MODELS FOR IOT

In IoT, data's are collected from various sources of sensors like temperature sensors, moisture sensors, etc. Thus, Data mining is used to extract or manage the data that are received from various IoT devices i.e. sensors. For IoT, Data Mining models can be categorized into three types. They are,

1. IoT in Multilayer Data mining Model
2. IoT in distributed Data mining model
3. IoT in Grid based data mining model

3.1 IoT in Multilayer Data Mining Model

The Multilayer data mining architecture is service oriented architecture. Multilayer data mining model consists of four layers, they are i) Data collection layer ii) Data Management layer iii) Event processing layer and iv) Data mining & Servicing layer. Data collection layer collects smart objects data by adopting the smart devices like RFID reader, sinks etc.

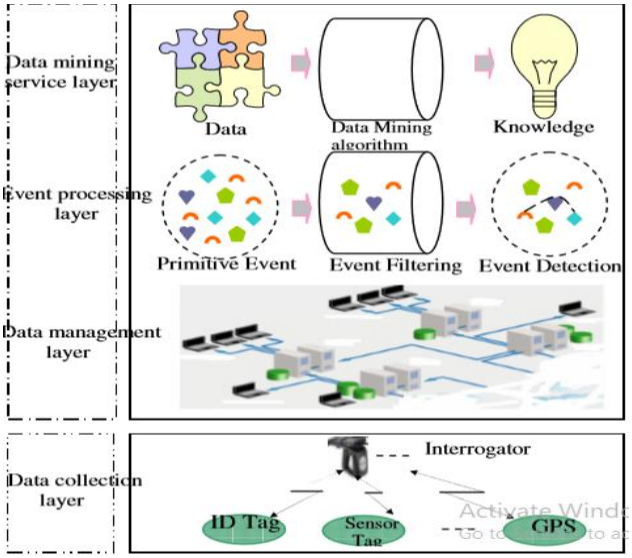

Fig 3. Multilayer data layer Architecture [1]

Data's are classified as data of RFID stream, GPS data, satellite data, sensor data, positional data etc. While collecting the data, the problems like energy efficiency, fault tolerance, filtration of data, communication of data etc are resolved. Data management layer uses centralized or distributive database to manage the collected data from smart objects.

In IoT, Smart objects are connected to each other with the help of this layer. Next layer is Event processing layer which analyzes the event in IoT where event is an integrated process various factors such as data, time etc. This integrated event helps to attain high level data processing mechanism which helps to aggregate, organize or analyze the primitive events after filtration. Last layer is called as Data Mining and servicing layer which is developed based on the previous two layers. Various data mining services such as classification, forecasting, clustering, outlier detection, association analysis etc which are object based or event based.

3.2 IoT in distributed Data mining model

In IoT, data has its own characteristics. IoT data's are mass, distributed, position related and time related. Also, IoT data's are heterogeneous. There are some problems addressed due to centralized data mining architecture. They are,

1. IoT data's are stored in different locations. So, it is difficult to mine the data from centralized database of distributed data.
2. Due to the large number of smart data, it requires pre processing and leads to the requirement of huge amount of central nodes.
3. Centralized architecture is not feasible because of data security, data privacy, fault tolerance, legal constraints and other factors.
4. Due to the limited resource nodes, it requires more energy and increases the transmission cost when the data's are sending to central nodes.

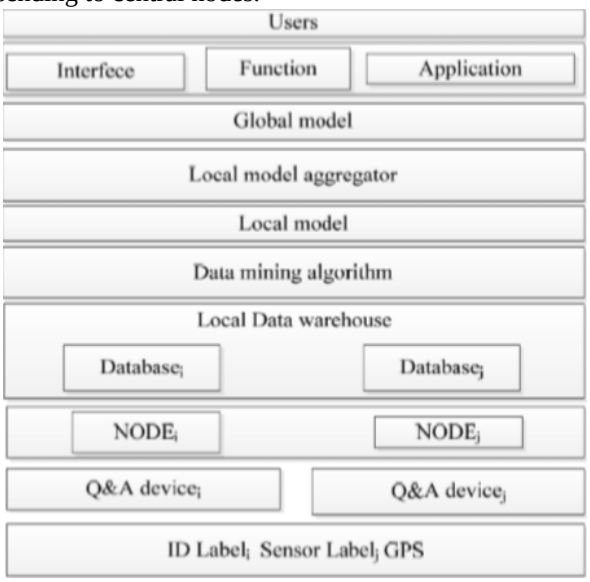

Fig. 4 Distributed data mining model in IoT [8]

Thus, Distributive data mining model reduces the requirement of high performance, high storage capacity and computing power. Distributed data mining model consists of two main layers namely, global and local node. In this architecture, there are multiple nodes in different locations /sites. Also, each node consists of local and global node. Data management should be done in both local and global nodes. Data mining activities are performed on local nodes first, and then it is available to all nodes by updating the global node.

3.3 IoT in Grid based data mining model

The main target or goal of IoT is to connect the various smart objects through internet. By this, it receives the high level context awareness, intelligent, interoperability, efficient smart objects. Thus, in grid computing, IoT plays a role of resource provider like device resources, data resources or computing resources. Grid miner is an infrastructure based technique which is used to extract the distributed data through distributed analytical processing. Grid Miner is building up on five layers in grid namely, resource layer, service layer, middleware layer, mining layer and application layer.

Also, Grid data mining of IoT uses the hardware like RFID, WSAN, WSN, Sensor nodes etc. Grid computing has the capability of efficient computing property over diverse environment. But, environment of IoT affords the diverse environment tie up in parallel.

Main merit of grid based data mining model is that it affords the services of parallel data mining process or parallel task execution. Also, it uses data
mining services of grid to implement data mining operations.

IV CHALLENGES OF DATA MINING OVER INTERNET OF THINGS & RESULTS

IoT data has been received from various diverse sources. Data mining is used in IoT to extract those data's. Due to the following challenges, the data mining process become more difficult.

- Rapid increase in large amount of data
- Diverse / Heterogeneous datasets
- Data Integration from diverse environment
- Requirement of real time analysis
- Security and Privacy

V CONCLUSION AND FUTURE SCOPE

In recent scenario, Data Mining and IoT both are an emerging and fast growing technology. In IoT, a large amount of data has been generated from various smart devices. To handle these data's, IoT needs integrated data and management. Also it requires useful data to be mined with the help of data mining techniques. In this paper, various data mining techniques like clustering, classification and association of data has been discussed to handle the diverse data and environment. Data mining models namely, Multilayer, Distributed and Grid based data mining models has been discussed with its architecture. In grid model, Grid Miner is a technique which affords distributed online analytics processing. In this paper, main challenges of data mining over IoT have been discussed.

Future scope is to identify the efficient data integration algorithm as well as IoT needs more security because it stores the data in cloud.

REFERENCES

1. Shen Bin, Liu Yuan, and Wang Xiayoi, Research on data mining models for the internet of things. In Image analysis and signal processing(IASP), 2010 International conference on pages 127-132. IEEE, 2010.
2. Tsai, Chun-Wei, et al. "Data mining for internet of things: A survey."Communications Surveys & Tutorials, IEEE 16.1 (2014): 77-97.
3. Stankovic, John. "Research directions for the internet of things." Internet of Things Journal, IEEE 1.1 (2014): 3-9.
4. Cooper J, James "A. Challenges for Database Management in the Internet of Things," IETE Tech Rev. 2009. 26:320-9.
5. Zhang, Chunguang, et al. "Analysis on Data Mining Model Objected to Internet of Things." International Journal of Advancements in Computing Technology 4.21 (2012).
6. M. Chen, S. Mao, Y. Zhang, and V. Leung, Big Data: Related Technologies, Challenges and Future Prospects, SpringerBriefs in Computer Science, Springer, 2014.
7. Feng Chen, Pan Deng, Jiafu Wan, Daqiang Zhang, Athanasios V. Vasilakos, Xiaohui Rong, "Data Mining For The Internet Of Things: Literature Review And Challenges", Hindawi, Ijdsn, 2015.
8. Divya Joshi, Chanchal Kumari, Abhishek Srivastava, "Challenges and Data Mining Model for Internet of Things", IJEAST, 1(3), Jan 2016.
9. G. Vamshi Krishna," Data Mining Processes, Applications, and Challenges for IoT", IJTRD, Volume 4(3), May – June 2017.
10. B. Yedukondalu, Dr. A. Daveedu Raju, "New Approaches of Data Mining for the Internet of Things with Systems; Literature Review and Compressive", IRJET, Volume 4(8), Aug 2017.
11. Asmita Gorave, Vrushali Kulkarni, "Discrimination Aware Data Mining In Internet Of Things", Ijca, 0975 – 8887, Volume 159, February 2017.
12. G. Kesavaraj and S. Sukumaran, "A study on classification techniques in data mining," in Proceedings of the 4th International Conference on Computing, Communications and Networking Technologies (ICCCNT '13), pp. 1–7, July 2013.
13. Furqan Alama, Rashid Mehmoodb, Iyad Katiba, Aiiad Albeshria, "Analysis Of Eight Data Mining Algorithms For Smarter Internet Of Things", Damis 2016.
14. Shweta Bhatia, Sweety Pate, "Analysis On Different Data Mining Techniques And Algorithms Used In Iot", Ijera, Vol. 5, Issue 11, November 2015.
15. Krushika Tapedia, Anurag Manohar Wagh, "Data Mining For Various Internets Of Things Applications", Ijrat, E-Issn: 2321-9637, Ncpci-2016.
16. Praveen Kumar, "Data Stream Clustering In Internet Of Things", SsrgIjcse, Volume 3, Issue 8, August 2016.
17. D. T. Larose, "k-nearest neighbor algorithm," in Discovering Knowledge in Data: An Introduction to Data Mining, pp.90–106, John Wiley & Sons, 2005.