© 2018 IJSRSET | Volume 4 | Issue 2 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
National Conference on Advanced Research Trends in Information and ing Technologies (NCARTICT-2018), Department of IT, L. D. College of Engineering, Ahmedabad, Gujarat, India In association with International Journal of Scientific Research in Science, Engineering and Technology

Application of Data Mining Techniques in IoT: A Short Review

Akshat Savaliya*[1], Aakash Bhatia[2], Jitendra Bhatia[3]
*[1]Computer Department, LD college of Engineering, Ahmedabad, Gujarat, India
[2]Computer Department, DA-IICT, Gandhinagar, Gujarat, India
[3]Computer Department, Vishwakarma Government Engineering College, Ahmedabad, Gujarat, India

ABSTRACT

Internet of Things (IoT) has been growing rapidly due to recent advancements in communications and sensor technologies. Interfacing an every object together through internet looks very difficult, but within a frame of time Internet of Things will drastically change our life. The enormous data captured by the Internet of Things (IoT) are considered of high business as well as social values and extracting hidden information from raw data, various data mining algorithm can be applied to IoT data. In this paper, We survey systematic review of various data mining models as well as its application in Internet of Thing (IoT) field along with its merits and demerits. At last, we discussed challenges in IoT.

Keywords: Internet of Things, Data Mining, Machine Learning, Application of Data Mining

I. INTRODUCTION

The Internet of Things(IoT) refers to the type of the network which connect anything i.e. physical objects-devices, buildings, vehicles and other items embedded with software, sensors and network connectivity based on stipulated protocols that enables these objects to collect and exchange data. In our daily lives, we have become more reliant on IoT with our wearable tech, appliances, our cars, how we receive health care. Due to Seamless integration of classical networks with IoT, it enables a great vision that all things can be easily monitored and controlled which results in to voluminous data. So, in order to make IoT more smarter, lots of data analysis is needed for which one of the most solution is data mining. Much research in recent years has focused on data mining in Internet of Things (IoT) which connects physical objects, person to person, person to machine or machine to machine via internet and manages information [11].

Data mining process refers to the process of semiautomatically analyzing large databases for pattern mining which are innovative, legitimate, useful and understandable which is also known as Knowledge Discovery in Databases (KDD). Data mining or KDD process includes problem formulation, data collection, data cleaning i.e. preprocessing, transformation, choosing mining task/method and result evaluation/visualization. Knowledge discovery is an iterative process.

Data mining overlaps with other fields like statistics, machine learning, artificial intelligence, databases but mainly it focuses on automation of handling large heterogeneous data, algorithm and scalability of number of features and instances.

As of late, an expanding number of rising applications manage countless sensor information in Internet of Things (IoT) due to a wide assortment of sensor gadgets on detecting layer. The extensive scaling of heterogeneous sensor creates a problem of data

handling which is one of key issue for the IoT framework application. Sensors in IoT applications sense the complicated environment and generates an enormous data that must be filtered and cleaned so that it can be interpreted and user will be provided with insights of the data collected in form of patterns [13]. Across various network infrastructures, IoT allows sensing of the objects and remotely access which in turn enables opportunities for a better integration between real and computerized world. It results into an improved effectiveness, exactness and better economic outcomes. Each query can be recognized exceptionally by the utilization of its installed registering framework, However these objects can interoperate inside the current infrastructure of the Internet. Assessments recommend that IoT will be an accumulation involving roughly 50 billion objects before the finish of 2020 [4]. The way toward finding and investigating helpful patterns in a large amount of information is what we refer it as Data Mining. Data mining can likewise be characterized as a sensible procedure that is used to investigate and look through extensive measure of enormous information so as to discover more valuable information in it. Till date the pattern finding procedures were not full fledge utilized and the information gathered was only an static accumulation of databases. But with the method of finding patterns in the information, more usage of the information is being acquired which settles on better choices for the advancement of the business or social aspect.

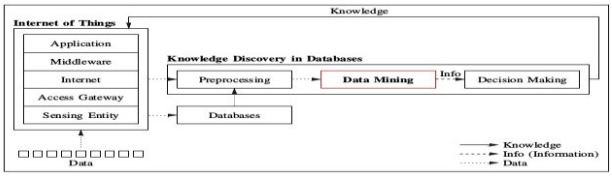

Figure 1. Data Mining Integrated IoT Architecture [13]

Figure 1 depicts clearly that, IoT gathers information from various sources, which may contain information for the IoT itself. KDD, when connected to IoT, will change over the information gathered by IoT into valuable data that would then be able to be changed over into learning. The information digging step is in charge of extracting designs from the yield of the information preparing step and after that encouraging them into the basic leadership step, which deals with changing its contribution to meaning information. It is essential to take note of that, every means of the KDD procedure may have a solid affect on the previous stage of mining. For instance, not all the traits of the information are valuable for mining; in this way, highlight determination is normally used to choose the key qualities from each record in the database for mining. The result is that information mining calculations may experience serious difficulties to discover valuable data (e.g., placing designs into suitable gatherings) if the chosen properties can't completely speak to the qualities of the information. It is likewise vital to take note of that the information combination, substantial scale information, information transmission, and decentralized processing issues may strongerly affect the framework execution and benefit the nature of IoT than KDD or information mining calculations alone may have on the customary applications. The key contribution of this paper includes:

- ✓ We explore the basic architecture of data mining assisted IoT.
- ✓ We discuss variants of data mining models for the IoTs.
- ✓ We survey various data mining application techniques used in IoT.

The rest of the paper includes the following. Section II discusses variants of work done in IoT using data mining techniques. Section III includes framework design of various data mining model. Section IV highlights various IoT application domain while section V discusses key issues involved in data mining of IoT. At the last, we conclude the paper.

II. RELATED WORKS

Since Internet of Things is a completely new concept, researches are still at the initial stage. Right now, there are few works regarding data mining in the IoT. Following are some of the work trending in this domain. Masciari [6] investigated mining in RFID data stream. Which tracks moving data made by different gadgets of IoT i.e. RFID sensor network, GPS gadgets, satellites and so on. Hector Gonzalez [5] proposed a model

utilizing which RFID data can be collected, which thus protect changes in it along with compression and path-dependent aggregation. Xiaolei Li [7] come up with a new system known as ROAM, which identify inconsistency in moving objects. SpatioTemporal Sensor Graphs (STSG) proposed by Betsy George [10] is used to model and mine sensor data. It can discover inconsistent patterns, incorporated areas at each time interim, and even hubs eligible to be future hotspot. Jae-Gil Lee [8] gave a new classification to track path followed by an object named TraClass using trajectory-based clustering and hierarchical region. Discovery of a knowledge from sensor data. Joydeep Ghosh [9] put forward a universal probabilistic system that permits supervised learning under computational/power/memory limitations. In the domain of data mining, a few broad organizations like Yahoo, Facebook, and Twitter pick up and supply attempts to open source ventures said by author in [10].

In [13], author proposed a design for high-performance data mining module of KDD for IoT with the three key considerations i.e. choosing objective, characteristics of data, and mining algorithm. Objective: The relevant mining techniques needs to be decided for the issue to be settled by the KDD. The suppositions, restrictions, and estimations of the issue should be determined first in order to accurately characterize the issue to be comprehended. With this data, the goal of the issue can be influenced precious stone to clear. Data: Another imperative worry of data mining is the characteristics of data, for example, size, distribution, and representation. Distinctive data typically should be processed in a different way. In spite of the fact that data originating from various issues might be alike, they may must be investigated distinctively if the implications of them are unique. Mining algorithm: Having above two parameters decided accurately, determining and selecting a data mining algorithm that suits to accomplish users task is very much easier task. In [13], author discussed three parameters which are very important to decide whether to develop new data mining algorithm or to use already designed algorithms. For an example, considering a scenario if we come to a derivation, that size and complexity of data that is needed to be processed is very high that are beyond available system capabilities to process and no other options or techniques are available to reduce size and complexity of data then it is supposed to be solved using novel mining algorithm.

III. VARIANTS OF DATA MINING MODEL FOR THE IoT

A. Multi-Layer Data Mining Model

As shown in Figure 2, model is partitioned into four layers namely information gathering layer, information administration layer, event processing layer and information mining service layer. Among them, information accumulation layer embraces devices , e.g. RFID sinks/readers and so forth., to gather intelligent information from different objects such as RFID stream information, GPS information, satellite information, positional information and sensor information and so on. Extraordinary sort of information requires distinctive information storage methodology. In the procedure of information collection, a progression of issues, e.g., vitality output, fault tolerance, data preprocessing, communications and so forth., ought to be very much explained [6]. Information i.e. data management layer focuses in centralized or disseminated database or information stockroom i.e. data warehouse to oversee gathered information.

An Event is a combination that joins information, time and different variables, so it gives an abnormal state component to information handling of IoT. An Event handling layer is utilized to investigate an events in IoT viably. Hence, it enables querying or investigation based on an event at this layer [10].

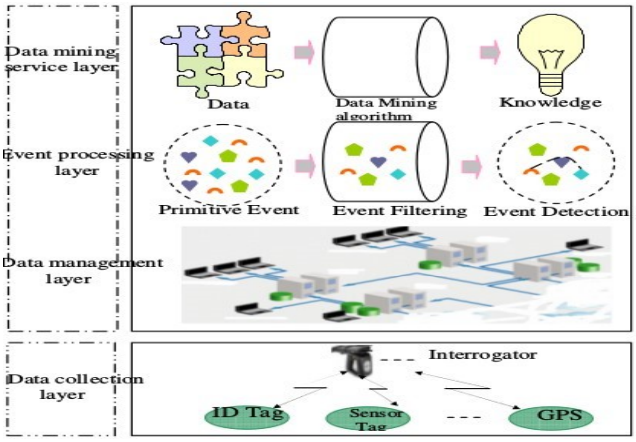

Figure 2. Multi Layer Model [1].

At that point, aggregation, sorting out and break down of information as per event can be done. Information mining administration layer is constructed in light of information administration and event handling. Different protest based or on the other hand management of event based information mining i.e. clustering, grouping, classification, forecasting, noise detection and mining of patterns, are provided for applications, e.g., SCM, inventory management and an optimization etc. The design of this layer is service-oriented.

B. Distributed data mining model

Comparing to traditional information which is raw in nature, Information in IoT has its own attributes. For instance, the information in IoT is dependably mass, appropriated, timerelated and position-related. At the same time, the information wellsprings of IoT are heterogeneous, and the assets of nodes are restricted. These attributes bring a few issues to unified information mining design. At initially, mass information of IoT is put away in distinctive locales. Consequently, it is troublesome for us to mine conveyed information by concentrated engineering. Furthermore, information in IoT is mass and needs preprocessing continuously.

For the thought of information security, information protection, adaptation to internal failure, business rivalry, legitimate requirements and different elements, the technique of assembling every pertinent datum is regularly not doable. In addition, the assets of nodes are restricted. The technique of transfering all information to central nodes does not enhance the utilization of vitality expensive transmissions. In most cases, the central node needn't bother with all information, yet a few evaluations of parameters. Hence, we can preprocess the raw information in the appropriated distributed node, and afterward send the fundamental information to the recipient.

Appropriated information digging model for IoT isn't just capable to take care of the issues brought by distributed capacity of nodes, but also decomposes problem complexity. In this way, the necessity of elite performance, high storage computing and processing power is decreased. In this model, the global control node is the center of the entire information mining framework. It picks the information mining calculation and the informational collections for mining, and afterward explores to the sub-nodes containing these informational collections. The sub-nodes get the raw information from different savvy objects. These raw information is supplied as an input to data filteration for preprocessing and then data abstraction and data compression, and finally, it get stored in the local data warehouse. Event separation, recognition and data mining at local nodes results into local models. Global models are the result of aggregation of local models are aggregated Subnodess trade protest information, process information and learning with each other. The entire procedure is restricted by the multi-agent based collaborative management module which is depicted in figure 3.

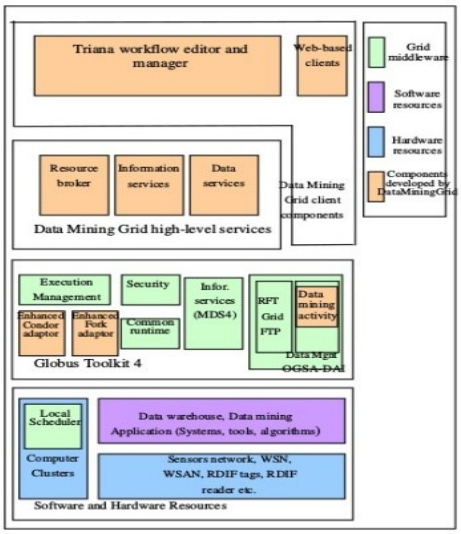

Figure 3. Distributed data mining model [1].

IV. APPLICATIONS

There are wide variety of application of data mining in Internet of Things. In [13] author suggested predicting user's preferences, nature and reaction to some situation, Object identification using different already available images of that object [13]. Video based classification where different scenarios and objects are identified, read facial expression of any person using already available devices like camera, microphone, etc, can also be done as suggested by author in [13]. Three-dimensional emotion model is used to identify human's emotions where a machine will have a very large data about a person different feeling and emotions that person posses while being in different situation and then deriving some patterns and conclude that persons emotions in particular given situation, Tracking movement of things sensing sound effects like human steps sound, door clapping, phone ringing, glass breaking [13].

Temperature, weather, wind speed, humidity prediction from previous data which may be very helpful to users like farmers or tourists before deciding their plans [2][16]. Also, Agriculture based on IoT, Cloud computing considered to be a great agricultural transformation [15]. Healthcare is booming domain for application of data mining using IoT devices and one can detect many deadly diseases in very early stage where getting rid of such disease is possible. Growth of disease in certain areas can be predicted using these techniques [13]. In academia domain, determining hot demanded areas according to student and market is easily achievable using data mining along with IoT devices [12]. Suggestions to deploy more public transport services in certain locations also can be evaluated using the data from IoT devices. Making prediction of usage of milk, grains, fruits, etc edible things in upcoming week or month or year can be derived from smart home IoT devices [3].

Utilization of IoT to gather information, which will be examined to get data valuable for basic leadership to enhance the web programs in Higher Education Institutions (HEIs) [5]. Many applications related to IoT essentially and implicitly consider occurrence of episodes (and events) with spatio-temporal constraints so as to initiate any further processing actions [9].

V. KEY ISSUES IN DATA MINING OF IOT

There are various issues involved in data mining in Internet of Things:

A. Efficiency in data gathering

Energy efficiency, scalability and fault tolerance should be taken into consideration when data is to be collected from distributed sensor networks [14]

B. Data abstraction and aggregation

Managing massive data generated from IoT is a challenging task. Efficient mechanism should be adopted for data deduplication.

C. Distributed data processing and mining

Due to nodes' constraints, paradigm shift is needed for prior level preprocessing of the data at each distributed nodes and an aggregated information is to be sent to sink node in order to optimize energy usage instead of sending all distributed data to server for processing.

D. Data mining towards the next age of Internet

In an upcoming generations of Internet, latest trends and technologies like ubiquitous computing, semantic web, IPv6 technologies are going to be integrated with IoT. This will give rise to challenges for Data Mining due to heterogeneous unstructure data [8].

VI. CONCLUSION

As a vital improvement of the next age of Internet, the Internet of Things pulls in numerous considerations by industry world and scholarly circles. IoT information has numerous qualities, for example, distributed storage, mass temporal and spatial related information, and constrained assets of nodes and so forth. These makes the issue of information mining in IoT turn into a test assignment.

VII. REFERENCES

[1] Shen Bin, Liu Yuan, and Wang Xiaoyi. 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] 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] 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.

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

[7] Rob Petersen. 20 companies do data mining and make their business better. 2016.

[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] 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.

[10] Aashi Singh and Shilpi Sharma. Analysis on data mining models for internet of things. In I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017 International Conference on, pages 94–100. IEEE, 2017.

[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] Fatima Merchant Suhem Parack, Zain Zahid. Application of data mining in educational databases for predicting academic trends and patterns. Technology Enhanced Education (ICTEE), 2012 IEEE International Conference, 2012.

[13] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, Laurence T Yang, et al. Data mining for internet of things: A survey. IEEE Communications Surveys and Tutorials, 16(1):77–97, 2014.

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

[15] Qiulan Wu, Yong Liang, Ying Li, and Yusheng Liang. Research on intelligent acquisition of smart agricultural big data. In Geoinformatics, 2017 25th International Conference on, pages 1–7. IEEE, 2017.

[16] 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.