Review on Integrating Data Mining with IoT: Applications and Future Challenges

P.Rathna Sekhar[1], J. Kalyani[2]
[1, 2] Dept. of Computer Science
[1, 2] University P.G. College, Osmania University, Secunderabad, Telangana, India.

Abstract- *Internet of Thing's (IOT) is a fast growing technology and everyday numbers of devices connected is increasing rapidly. By the end of 2020 the number of connected devices to IoT will reach 5 billion and revenue generated by IoT will be 3 trillion dollars. Internet of Things (IoT) generates very crucial and precise massive dynamic data every day. The prime aspect of IoT is to push the users towards smart interactions like automation of tasks, and automatic decision making etc. Integration of data mining techniques with IoT allows effective processing of this data and get desired useful information. The successfulness of IoT is depends on how effectly it carry outs data mining process to support users in smart decision making. The objective is paper is to provide brief review about IoT, data mining techniques used in IoT, and various applications of data mining and IoT integration. Finally, it discusses future trends and challenges of IoT data mining field with the help of secondary data collected from various journals, articles research findings, and citations.*

Keywords- Data Mining, IoT, Integartion of Data Mining with IoT, Applictions, Future Challenges.

I. INTRODUCTION

Internet of Things (IoT) is a technology that connects various computing and communication devices and offers advanced services. The primary intention of IoT is to automate the smart decision making by providing real-time and historic data. Its applications will enable us to mange and predict various day to day activities such as managing health, smart cities, and forests etc, and predicting weather, traffic, crowd and energy use and emergency situations etc. At present there are more than 1.5 billion devices are connected trough IoT. A recent analysis from Statista (Figure 1) states that the number of connected Internet of Things (IoT) devices in the world from 2015 to 2025 will be more than 75 billion. By 2020, the total number of Internet of Things devices will be progressed to 31 billion. And the overall market value will be reached to 3 trillion dollars[5][2][24]. IoT is illustrated as uninterrupted connection of data and things and in IoT everything is same treated as things. These things may be physical things or virtual things and are connected to each other irrespective of location and time. These may compromise users, vehicles, data centers, wristwatches, household equipments, computing and communication devices like Computers, Laptops, Smartphones, chips and sensors, Bluetooth, Infrared, cellular networks, Near Field Communication (NFC), Wi-Fi networks, RFID and fitness tracking devices etc.

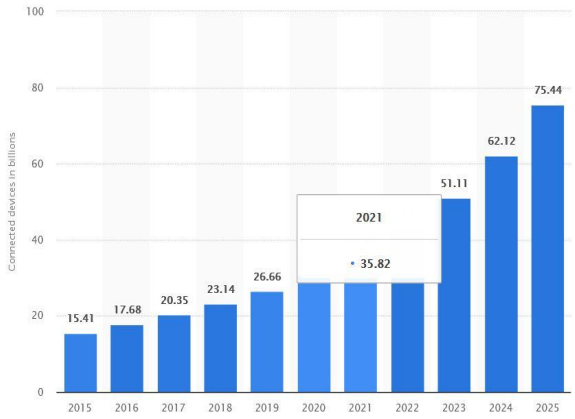

Figure 1. IoT Number of Connected Devices in the World from 2015 to 2025 (Source: Statista)

These IoT devices generate very huge volume of complex, rough and dynamic data. It is difficult task to conserve, process that data and produce effective business knowledge. For providing that kinds of services to improve of the processing speed and produce accurate results data mining is essential. Data mining or Knowledge Data Discovery (KDD) is a process of getting effective and productive or meaningful information from a very large set of data with the help of relationship between the data or patterns[23][9]. Data mining compromises of finding compelling, innovative, interesting, and possibly needful patterns from immerse data sets and applying Data mining techniques on those data sets for extraction of hidden and required information. In other words data mining is a process of knowledge extraction, discovering knowledge in databases (KDD), data/pattern analysis, information harvesting, data archeology and data dredging, etc[5][16]. Data mining process includes steps Data preparation, Data mining, Data presentation. Data preparation is a process of collecting data from various data sources, removing the noise from collected data, and preprocessing of data to make it useful for data mining step. Data mining step

applies techniques on the data to discover meaningful patterns and analyze those patterns for discovery of knowledge. Data presentation visualizes the data and produces the discovered knowledge. Most significant thing is, these steps have strong impact on the end results of mining process.

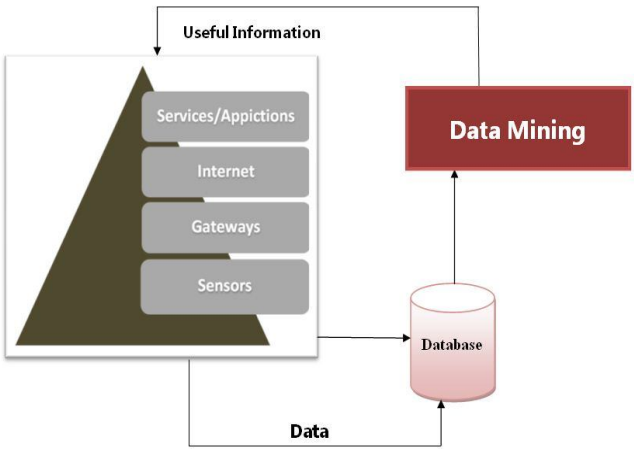

Figure 2. Integration of Data Mining with IoT.

IoT collects data from various sources i.e. things. This data will be used to understand, manage and control complex environments, enables greater automation, highly efficient, productive, accurate, and precise decision making knowledge. With data mining techniques applied on data collected by IoT (Fig. 2), useful and functional information will be generated, after that this information is converted into user required knowledge. Only important attributes of data from the entire records are selected for knowledge extraction. It is difficult to get desired information, without complete incorporation of required attributes. Data amalgamation, large volumes of data, computing issues and data transmission, will have powerful influence on the performance and service of IoT. The following sections discuss about IoT and data mining applications, and future prospective of integrating data mining process with IoT environments.

II. DATA MINING PROCESSES TECHNIQUES FOT IOT

A. Clustering

Dividing data into significant, relevant and meaningful groups of patterns on the basis of specific features and attributes, in which each group has similar features and related patterns and different groups have unrelated and dissimilar groups is called clustering[8][24]. Principally, it is an unsupervised learning technique that catalogues the data into sensible and meaningful groups. Numerous TV Shows, Movies, YouTube videos can be into categorized into groups according to the content present in those videos etc, which will helps users to watch desired shows, movies or videos based on their interests[23]. Since clustering is an unsupervised technique we apply this technique without any deep knowledge about the data. In IoT applications, sensors collect unstructured kind of data such as user behviours, weather reports, traffic information, clustering can be applied that collected data for finding useful patterns which helps us to predict user behviour, fitness and weather forecast, etc. For example, by observing this data collected by IoT activity tracker sensors, we can find out the patterns in that are in accord with activity score at a certain time and can compare the scores in the data. Now we can group the activities on the basis of the activity scores at a certain times. We can also group the timings of the grouped activity scores. This technique is called clustering. Several researchers developed various clustering algorithms such as k-Means, hierarchical, k-Medoids and Density-based[3], for enhanced and refined outcomes in IoT. The efficiency and correctness along with applying clustering in suitable situations and on accurate problems should be primary objectives for the IoT. It is used to discover the relationship between the users and IoT nodes. The end results of clustering are used to predict the traces of IoT nodes and provide knowledge to the users.

B. Classifiction

Classification technique is crucial for decision making process. The goal of classification is to accurately predict the desired class for each case in the dataset. Classification is a technique that labels data items. With help of this technique one can easily find out the category of certain item in a dataset. For example, classification technique we can effectively analyze and predict user buying behavior by user's family background, age, and profile etc. Classification technique is used at the time of the collected data is in correspondence with kinds of classes. This technique is only classifies data but not useful in prediction of the data. For example IoT activity tracking sensor's data is linked with different activities performed by users. During the end of the day classification can be used to track overall day to day activity. The widely used algorithms of classification technique for IoT are Decision tree, k-nearest neighbor algorithm, support vector machine, and ariori algorithm. Decision tree algorithm is tree structured algorithm which looks like flowchart. Based on the feature values, decision tree classifies objects. The apriori algorithm is a process of calculating the frequency of itemsets by scanning the database. The K-Nearest neighbor (KNN) algorithm is used to discover the closest point of the detected object. The primary objective of KNN is to detect the K nearest points. Bayesian networks are directed acyclic graphs in which nodes illustrate random values and edges represents conditional dependencies of those

values, to each other. Support Vector Machines algorithm (SVM) is supervised learning algorithm with correlated learning algorithms which analyzes the data and finds out patterns, according to statistical learning. These classification algorithms can be applied on various kinds of data sets and based on the performance of these algorithms they can be used to predict the natural disasters for example abnormal weather changes such as cyclones, earth quakes, etc.

C. Feature Selection and Time Series Analysis

Feature Selection is a key process of data mining that allows us to select the attributes of data which influence certain event. For example diet, work time and regular sleep times are selected as attributes to guide us to have efficient sleep time. Time series analysis estimates approximate calculations of future data that can be extracted from a certain extracted and analyzed dataset. This data has large size, high dimensionality, and dynamically updates at regular time periods. Applications of this algorithm include weather forecasting and finding density of air etc.

III. APPLICATIONS AND FUTURE CHALLENGES OF INTEGRATING DATA MINING WITH IOT

A. Applications

Integration of Data mining with has broad range of applications such as prediction of user preferences, behavior and reactions at certain situations. Expeditious growth of IoT resulted significant set of applications with data mining integration. These include automated day to day activities, smart city based applications like community health care, traffic management, electronic governance, disaster management, driverless vehicle driving, navigation energy efficient industry developments and weather forecasting classifying various situations based on video data available and identifying and recognizing objects, reading facial expression of persons, temperature and weather forecasting, areometry, predicting humidity from previous data etc. Applying relevant data mining techniques on IoT systems in smart cities enables existing system smarter and finer. At time of traveling, several IoT devices like GPS navigation systems, mobile phones, and smart sensors deployed in vehicles acts as data points and enables in predicting incidents heavy traffic, accidental zones and construction places etc. Knowledge about the reason for traffic congestions is publicly known with these data points. By applying classification algorithm we can overcome the traffic blockage and disperse the traffic. Certain areas are to be classified on the basis of possibly that may have higher, lower or medium occurrences of traffic congestion incidents. We can also predict the time, when will the blockage occur and avoid that route, and we can choose route with no traffic blockage to arrive the destination in lesser time. Replacing of conventional electric meters with smart electronic meters enables us to track real time energy consumption information. To collect energy consumption data automatically, and to detect anomaly in energy consumption, to predict and save energy consumption, Time Series analysis is used. For Governments IoT provides several smart city applications such as smart power and lighting management, smart parking meter, smart surveillance, smart traffic management, effective disaster management, productive resource management and adaptive emergency response system, Smart transportation applications such as roadways, railways, transport, port and aviation management, Smart grid, Smart infrastructure applications and Environment based applications such as power efficiency, Smart water applications such as waste water management and domestic waterworks, improving air quality, monitoring environment, land filling and waste management[23]. On the industry IoT offers services, smart energy saving, smart healthcare, industrial IoT, asset management, smart inventory control and smart customer relationship management, smart agriculture applications, Smart banking applications, smart building and smart construction, and smart education, smart manufacturing, and smart logistics applications[24].

B. Future Challenges

Integrating IoT with data mining provides variety services to every individual, governments, industries and organizations. Every day more number of devices are connecting because of its vast range of applications. With integration of data mining, IoT can produce more precise and smart applications. In a recent report of Compound Annual Growth Rate (CAGR), it is stated that IoT along with data management market is expected to increase from 27.54 Billion USD in 2017 to 66.44 Billion USD at the end of 2022. IoT data management includes Data Analytics, Data Security, Data Integration, Data Visualization, Metadata Management, Data Services and its market will see a global growth of 19.3% by 2022. In future, IoT makes living, manufacturing, agriculture, and health domains smarter. With revolutionization of Industry 4.0 IoT data analytics or mining plays crucial part. IIoT (Industrial IoT) is most happening technology revolution right now. IoT Data mining has several limitless applications like green energy, reducing energy consumption by automatically supervised buildings, smart traffic management etc, but implementation of these will be a big challenge. With IoT, security is most vital concern since number of connected devices increasing rapidly. The crucial problems in security are potential for unexpected uses of consumer data, ubiquitous

data collection, and severe security risks. With restrictions in accessing data, consumers feel difficult sometimes. IoT has some sort of complex shared standards and infrastructures. Geographical communication of sensors, hardware, applications and devices should be enabled. According to Feng Chen[5], mining of data is one of the major challenges facing by the IoT. Fortunately, new data mining techniques making it easier and the advancement of the IoT would be the next phase of the big data revolution. And Management of Big Data will become burdensome and complex if proper algorithms nit applied.

IV. CONCLUSION

With rapid growth of IoT new data will be added everyday at large amounts. To deal with large data and for getting desired knowledge from IoT systems data mining is needed. In this paper reviewed data mining techniques, applications on IoT. This per also discussed various future challenges in this prospective. Inspecting the vast amount of data and finding useful knowledge for future prospective is a big task with fast paced growth of IoT, big data, and cloud computing. Data analysis became more complex with IoT, and data is collected from various kinds of sources in different forms like video, audio, images, SQL tables, CSV sheets, and web documents. Before applying big data analytics that data must be processed, cleaned and converted into a single understandable structure. Some applications need real time data extraction and processing. If real time data is huge volumes this would become a big challenge.

REFERENCES

[1] Jihyun Song, Kyeongjoo Kim, Minsoo Lee, "A Big Data Analysis and Mining Approach for IoT Big Data", ijacst, Volume 7 No.1, January 2018.

[2] Furqan Alam, Rashid Mehmood, Iyad Katiba, Aiiad Albeshri, "Analysis of Eight Data Mining Algorithms for Smarter Internet of Things", DaMIS 2016.

[3] Shweta Bhatia, Sweety Patel, "Analysis on different Data mining Techniques and algorithms used in IOT", ijera, Vol. 5, Issue 11, November 2015.

[4] Akshat Savaliya, Aakash Bhatia, Jitendra Bhatia, "Application of Data Mining Techniques in IoT: A Short Review", IJSRSET, NCARTICT-2018.

[5] Feng Chen, Pan Deng, Jiafu Wan, Daqiang Zhang, Athanasios V. Vasilakos, Xiaohui Rong, "Data Mining for the Internet of Things: Literature Review and Challenges".

[6] http://blog.buildinginternetofthings.com/2012/10/13/data miningandtheiot/.

[7] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang, "Data Mining for Internet of Things: A Survey", IEEE Communications Surveys & Tutorials, Vol. 16, No. 1, First Quarter 2014.

[8] Feng Chen, Pan Deng, JiafuWan, Daqiang Zhang, Athanasios V. Vasilakos, and Xiaohui Rong, "Data Mining for the Internet of Things: Literature Review and Challenges", Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, Volume 2015.

[9] Krushika Tapedia, Anurag Manohar Wagh, "Data Mining for Various Internets of Things Applications", NCPCI-2016.

[10] Mehmet BEYAZ, "Data Mining Techniques", TTG International, L.T.D, June 06/2016.

[11] Praveen Kumar, "Data Stream Clustering in Internet of Things", SSRG-IJCSE, Volume 3, Issue 8, August 2016.

[12] Asmita Gorave, Vrushali Kulkarni, "Discrimination Aware Data Mining in Internet of Things (IoT)", IJCA, Volume 159, Issue 3, February 2017.

[13] Thirunavukarasu B, Dr T.Kalaikumaran, Dr S.Karthik, "INTEGRATION OF DATA MINING AND INTERNET OF THINGS – IMPROVED ATHLETE PERFORMANCE AND HEALTH CARE SYSTEM", ijtra, Special Issue 11, Nov-Dec 2014.

[14] Dale Benton, "IoT and data analytics: the mining of tomorrow", Feb 2018, www.miningglobal.com.

[15] Albert Bifet, "Mining Internet of Things (IoT) Big Data Streams", Universit´e Paris-Saclay.

[16] Aryan Singh, "Processing and Mining Data in IoT Systems and Enterprise Applications", GlobalLogic Inc.

[17] Annalisa Appice, Michelangelo Ceci and Donato Malerba, "Relational Data Mining in the Era of Big Data".

[18] Yan Chen, Ai-xia Han Cai-hua Zhang, "Research on Data Mining Model in the Internet of Things", AMCCE 2015.

[19] Shen Bin, Liu Yuan, Wang Xiaoyi, "Research on Data Mining Models for the Internet of Things", Zhejiang University.

[20] 20 Ryan Kh., "The Biggest Data Mining Challenges Facing IoT", Feb 2017, https://dzone.com.

[21] Harbor Research, "Smart Systems and IoT Analytics".

[22] Harley Oliff, Ying Liu, "Towards Industry 4.0 Utilizing Data-Mining Techniques: a Case Study on Quality Improvement", CIRP 63, 2017.

[23] G. Vamshi Krishna, "Data Mining Processes, Applications and Challenges for IoT", IJTRD, Volume 4, May-June 2017.

[24] Sallauddin Mohmmad, G. Sunil, "A SURVEY ON NEW APPROACHES OF INTERNET OF THINGS DATA MINING", IJARCS, Volume 8, No. 8, September-October 2017.