2nd International Conference on Recent Developments in Science, Humanities & Management
Mahratta Chamber of Commerce, Industries and Agriculture, Pune (India)
8th July 2018 www.conferenceworld.in ISBN : 978-93-87793-33-0

INTERNET OF THINGS USING DATA MINING: CHALLENGES AND APPLICATIONS

Mrs. Sandhya Gundre [1], Ms. Ketaki Bhoyar[2] ,
Mrs. P. P. Shevatekar[3], Mrs. Nalini Yadav[4]

[1,2,3,4]*Department of Computer Engg.,DYPIEMR, Pune, India.*

ABSTRACT

The Internet of Things (IoT) visualizes a future in which everything on the earth digital and physical objects (e.g., smart phones, TVs, cars) can be connected together via internet. The IoT's characteristics, including a global network infrastructure with self configuring capabilities of things and device with network level heterogeneity and intelligent interface, all will make development of the diverse applications and services, a very challenging task.

Initiaaly, mining information and knowledge from large databases that is generated or captured by IoT, has been recognized very useful for many applications. secondly, Data mining is playing a critical role in making smart system to provide more convenient services and environments, Data Min ing also provide too many researchers as a key research topic that are associated with the development of IOT enabling technologies, infrastructure and applications. This survey begins with a introduction of Internet of things. And a brief review of the "role of data mining in IoT", in which we discussed a Data mining model for IOT to provide a smart world. In addition challenges, applications and future trends of this field are highlighted.

Keywords : Distributed processing, Internet of Things (IoT) applications, data mining, knowledge discovery

1. INTRODUCTION

The Internet of Things (IoT) refers to future generation of Internet which will include smart environment of uniquely identifiable object and self aware things to handle large web servers and supercomputer clusters and build a new smart world [1][32][36].

This future internet will create the new revolution in the field of information technology after the revolution of

computer and Internet. The basic idea of this concept is the pervasive presence around us of a variety of things or objects which are capable to interact with each other to perform a specific task and manageable by their own in case of failure.[2][29].

.

S. Haller et al.[3] have given the following definition of IoT : "A world where physical objects are seamlessly integrated into the information network, and where the physical objects can become active participants in business process. Services are available to interact with these 'smart object' over the Internet, query their state and any information associated with them, taking into account security and privacy issues."

IOT is not only a technology but it is a concept ,which is applicable to smallest area such as home ,street light signal to a wider area such as real time application to support ,assets and providing new services. These are providing lot of business opportunities but also adding complexity to it.

The internet of thing are creating large amount of the data and it will be create more data in future also.So ,it is necessary to developed methods to effectively handle the data, analyze it and mining it.IoT data can be categorized into several types as descriptive data, environmental data ,positional data RFID data stream and sensor network data[4]. It brings the great challenges for managing , analyzing and mining different types of IOT data.

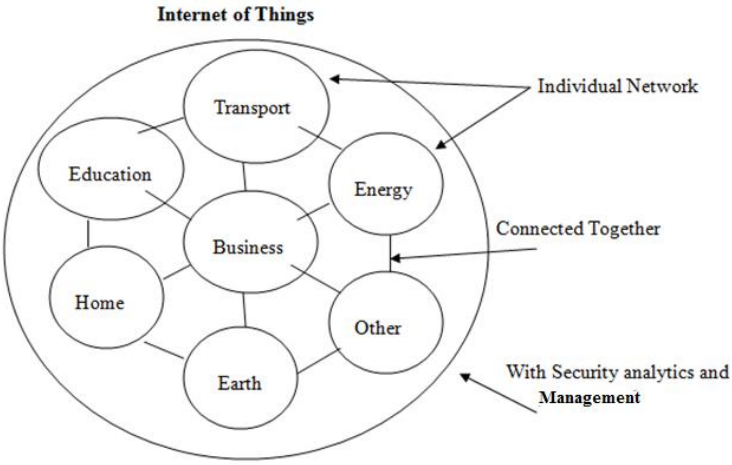

Fig 1 : IOT viewed as a network of networks

Recently, the large amount of data collected and stored around the world has been increasing at the exponential rate. IOT produced data are heterogeneous and large scale, because it is produced by different type of IoT devices like sensors, networks, the web, social media and transactional devices. the traditional methods and technologies are not suitable for storing and analyzing the complex and large scale data [39]. New technologies have boosted the ability to store, process and analyze the massive data.[39] For example Cloud-based Technologies and platforms gain in popularity, new data mining and machine learning algorithms have been developed as the Cloud services. Moreover, another new trend known as big data brings new challenges to data mining due to large volumes and different varieties of data. Thus, the Cloud and big data not only yield new data storage and processing mechanisms but also introduce ways of the intelligent data analysis[39]. Due to the complexity of data, various data mining Methods should be used jointly. The goal to make massive data mining

simpler, effective, and easily understandable for data scientists and business intelligence professionals can be achieved by constructing scientific workflows for data mining process using a drag and drop interface. The usage of scientific work- flows allows composing the convenient model of data mining process covering a number of different methods. Thus, the simulation and solving of real-world time- and resource-consuming data mining problems may be realized.

This paper gives a overview of Internet of things of data mining with a detailed discussion about how to connect it to the IoT to shift to the next generation internet environment. The focus in III section is to develop effective methods or models for managing, analyzing and mining data. The challenges of data mining toward IOT technologies for IoT are presented in Section IV. Conclusions and future trends are drawn in Section V.

2. RELATED WORK

Internet of thing has become part of your daily life, however the internet of thing is still maturing "The Internet of Things allows people and things to be connected anytime, anyplace, with anything and anyone, ideally using any path/network, and any service" [5], [6] as definition given by European research cluster of IoT (IERC) The International Telecommunication Union (ITU) views IoT very similarly: "From anytime, anyplace connectivity for anyone, we will now have connectivity for anything" [7].

The ERC vision is that "the major objective for IOT is the creation of smart environment and self aware thing (includes a variety of physical elements). These include for example: smart transport, products, cities, buildings, rural areas, energy, health, living, etc, as smart phones, tablets, and digital cameras. Based on this view of "things," an enormous number of devices will be connected to the Internet, each providing data and information, and some, even services[32].

Several services and applications of different technology levels coexist within the current utility grid. In this sense, it is necessary to establish techniques that provide the capability to integrate information from different architecture and technological levels. These technologies increase the robustness of the management systems related to the utility grid.

Basically, Data mining functionalities include classification, clustering, association analysis, time series analysis, outlier analysis, etc.

- Classification is the process of finding a set of models or functions that describe and distinguish data classes or concepts, for the purpose of predicting the class of objects whose class label is unknown.
- Clustering analyzes data objects without consulting a known class model.
- Association analysis is the discovery of association rules displaying attribute-value conditions that frequently occur together in a given set of data.
- Time series analysis comprises methods and techniques for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.
- Outlier analysis describe and model regularities or trends for objects whose behavior changes over time [12].

3. DATA MINING MODEL FOR THE INTERNET OF THINGS IoT

The first It is predictable that future software industry will produce large amount of data (or big data) from various IoT based services and applications [21]. To store and analyze these large scale and heterogeneous data, and filtered information from these data has become an important task of these smart organizations because this task can put them in an unassailable position (i.e., a position to provide unique services) or it will create for them much more commercial values (i.e., innovation of products)[14]. Two majorly defined category of data, first is "data about thing" and another one is "data generated by things", the first describe the data related to things that is used to maximize the system performance, infrastructures, and things of IoT and second describe the data that is generated by things that are the results of interaction between humans, between human and machine, and between machines. This generated data is required further classification approach to improve the services provided by IoT [13][14].

As depicted in Fig. 2, IoT collects data from various sources, which may contain data for the IoT itself. KDD, when applied to IoT, will convert the data collected by IoT into useful information that can then be changed into knowledge. The data mining step is accountable for extracting patterns from the output of the data processing step and then feeding them into the decision making step, which takes care of transforming its input into useful knowledge. It is important to note that all the steps of the KDD process may have a strong impact on the final results of mining. For example, not all the attributes of the data are useful for mining; so, feature selection is usually used to select the key attributes from each record in the database for mining. The consequence is that data mining algorithms may have a hard time to find useful information (e.g., putting patterns into appropriate groups) if the selected attributes cannot fully represent the characteristics of the 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 [24].

As shown in figure 2,IoT collects data from various sources like Application, internet, middeleware ,iot itself too.

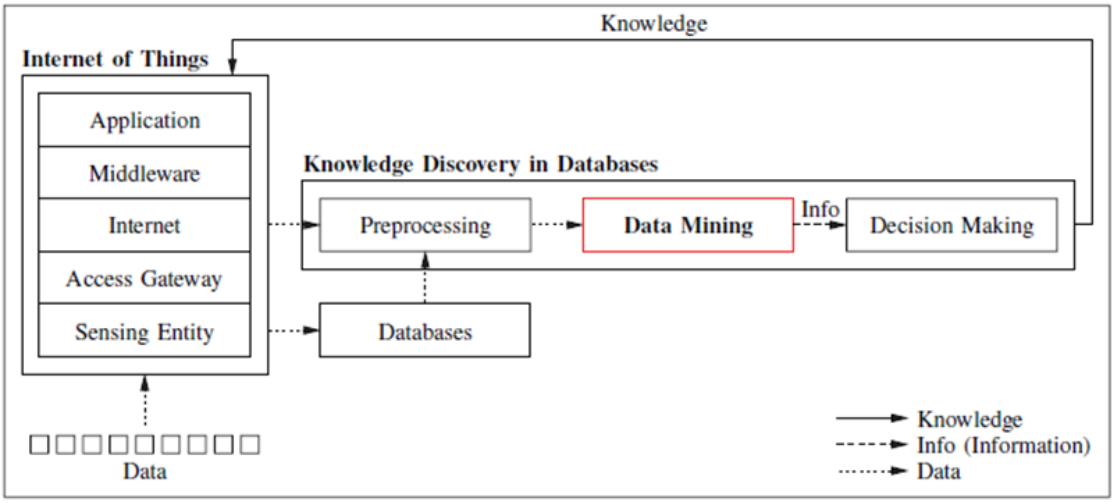

Figure 2 : The architecture of IoT with KDD.

3.1 Multi layer data mining model

According to the survey of big data mining system and IoT architecture, we suggest the multilayer architecture for IoT and big data mining system. In this system, it includes 5 layers software and hardware resource layer, Data collection layer, data management layer, data processing layer and service layer as shown in Figure 7

FIGURE 7: The suggested big data mining system

- software and hardware resource layer : its include various software and hardware IoT devices to collect various smart object's data, which are RFID stream data, GPS data, satellite data, positional data and sensor data etc, such as sensors, RFID, cameras and other devices.
- Data Collection layer: In the big data mining system, structured data, semi-structured data and unstructured data can be integrated.
- Data Management Layer: This layer applies centralized or distributed database or data warehouse to manage collected data. After object identification, data abstraction and compression, various data are saved in the corresponding database or data warehouse. Real-time data and batch data can be supported and all data can be parsed analyzed and merged [25].
- Data processing:this layer integrates the data, time and other factors, so it provides a high-level mechanism for data processing of IoT [37]. Lots of open source solutions are integrated, including Hadoop, HDFS, Storm, Oozie,etc. data processing layer is used to analyze data in IoT effectively. Then we can aggregate, organize and analyze data according to the requirement [25].
- Service: Data mining service layer is depends on data management and data processing. Data mining functions will be providing as service. Various object-based or event-based data mining services, such as classification, forecasting, clustering, outlier detection, association analysis or patterns mining, are provided for applications, e.g., supply chain management, inventory management and optimization etc[25].
- Security/Privacy/Standard: Security, privacy and standard are very important to big data mining system. Security and privacy protect the data from unauthorized access and privacy disclosure. Big data mining system standard make data integration, sharing and mining more open to third part of developer[25][26].

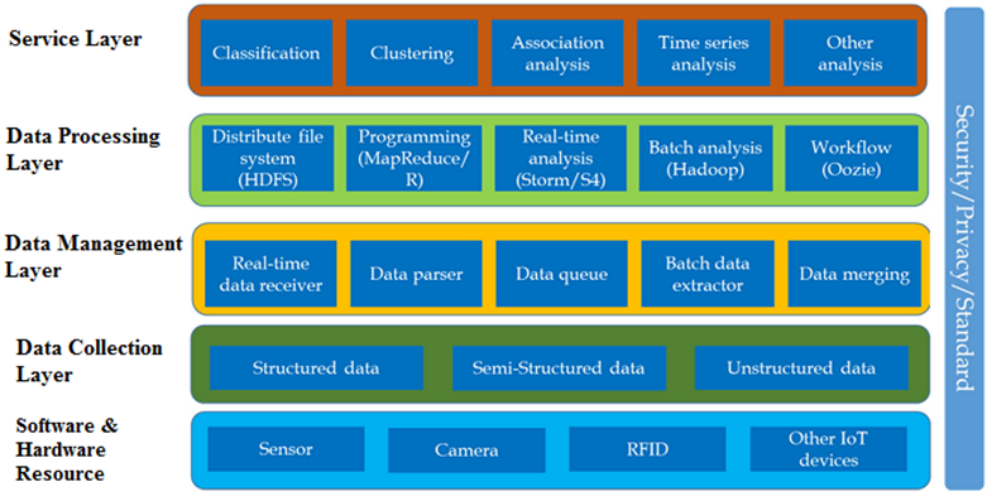

Figure 3 : Multi-layer data mining model for IoT

4. CHALLENGES AND OPEN RESEARCH ISSUES IN IOT AND BIG DATA ERA

4.1. Data collection and mining information from smart objects of IoT: The first challenge is to access, extracting large scale data from available local and wide-area computer network, including Internet also from huge distributed, heterogeneous databases [27][28] We need to mining knowledge from different types of formatted or unformatted data with the variety, heterogeneity and noise of the data, and it's a big challenge to find the fault and even harder to correct the data[26][27]. In data mining algorithms area, how to modify existing algorithms to big data environment is also another big challenge.

4.2. Data Classification and analyzing

The discovered knowledge should accurately describe the contents of the database and be useful for certain applications[27]. In big data era, the data sources are diverse, for example, we need to integrate sensors data, cameras data, social media data and so on, all these data are different format, byte, binary, string, number, etc. Noise and exceptional data and uncertainty of data should be handled elegantly in data mining systems. Data Classification and analyzing meaning is developing a effective methods should be developed for filtering redundant data. This also motivates a systematic study of measuring the quality of the discovered knowledge, including interestingness and reliability, by construction of statistical, analytical, and simulative models and tools [25][27].

4.3. Data Mining Algorithm for IOT : today's ,most data mining algorithm work with vector valued data.it is an important challenge to extend data mining algorithm to work with other data types, including time series and process data, unstructured data ,semi structured data ,multimedia and collaborative data, hierarchical and multiset data and collection value data[12]. Also most data mining algorithm require bringing all together data to be mined in a single , centralized data warehouse.

A fundamental challenge is to developed distributed versions of data mining algorithm so that data mining can be done while leaving some of the data in place[37]. The main tasks include classification, forecasting, clustering, outlier detection, and association analysis, spatial and temporal patterns mining for IoT [28].

Thus, the design of privacy preserving schemes with great data utility remains great challenges on data privacy preservation in IoT for future research.

The main objective of big data mining model is to develop an efficient structure to handle vast amount of data that is generated by IOTs and mining that data to extract useful information. Following key point we need to be consider during design of data mining framework, security & privacy of data, the data sharing mechanism, the growth rate of data , etc. To create a well designed data mining framework for IOT is a very important direction and a big challenge.

5. DATA MINING MODEL TOWARDS THE NEXT GENERATION OF INTERNET:

Data mining today is at best a semi automated process and perhaps destined to always remain so. On the other hand a fundamental challenges is to develop data mining system which are easier to use[12] and able to sustain with all next generation environment of internet. The next generation internet (NGI)will connect sites at OC-3(155 MBits/Sec),speed and higher. This is over 100 times faster than the connectivity provided by current networks. With this type of connectivity, it becomes possible to correlate distributd datasets using current algorithms and technique. In addition new protocol, algorithm and languages are being developed to facilitate distributed data mining using current and next generation networks.

In addition ,the development of data mining and knowledge discovery environment that address the process of collecting ,processing, mining and visualizing data as well as the collaborative and reporting aspect necessary when working with data and information derived from it, is another important fundamental challenge[20][25].

6. IOT AND DATA MINING APPLICATION

The internet of things is the name given to future of connected devices. there are two clear subsets, the “consumer IOT” that includes wearable computers, smart household devices and network appliances, another one is “Industrial IOT” includes networked smart grid, manufacturing, medical and transportation [31].Some IOT application are addressing societal needs and advancement to enabling technologies such as nano electronics and cyber physical systems continue to be challenged by a variety of technical, institutional and economical issues[38]. Here are some of the IoT smart application [34] discusses in this section.

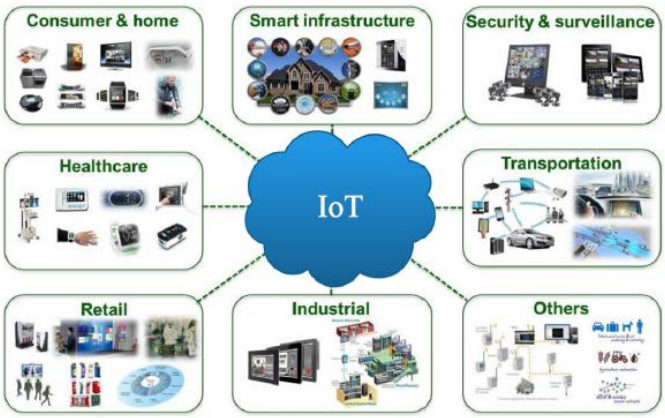

Figure 4 : IOT Applications

7. CONCLUSION

To discover useful knowledge from real-world and simulated big data, business intelligence professionals and data scientists face with new challenges. In today smart era, the conventional technologies and methods cannot store and analyze a large amount of data. Since the "Internet of things" gain in popularity, the attention is focused on the development of new data mining model for internet of things with a possibility to access high-performance computing environments. In This paper, review of the existing data mining software and solutions with internet of thing, has been drawn. Smart environment features and the new technologies related to information management are the future of the new smart services and applications. Several services and applications of different technological levels coexist within the current utility grid. In this sense, it is necessary to establish techniques that provide the capability to integrate information from different architecture and technological levels. These technologies increase the robustness of the management systems related to the smart environment. We also discuss the challenges and issues of data mining towards IOT, and various applications of internet of things. We believe that further research is needed to construct efficient models for such applications the models could be improved by adding new information, and performing the modeling algorithm. Future research lines include the application of this technology to other types of database, such as document-based and key-value databases.

REFERENCES

Journal Papers:

[1] Anne James, Joshua Cooper, Keith Jeffery, and Gunter Saake "Research Directions in Database Architectures for the Internet of Things: A Communication of the First International Workshop on Database Architectures for the internet of things (DAIT 2009),"BNCOD 2009: 225-233.

[2] 1. Luigi A., Antonio I., Giacomo M. 2010.The Internet of Things: A survey. Science Direct Journal of Computer Networks, Volume 54, Pages: 2787–2805.

[3] S. Haller, S. Karnouskos, and C. Schroth, "The Internet of Things in an enterprise context," Future Internet Systems (FIS) , LCNS, vol. 5468. Springer, 2008, pp. 14-8.

[4] Cooper J, James "A. Challenges for Database Management in the Internet of Things," IETE Tech Rev. 2009. 26:320-9.

[5] P. F. Harald Sundmaeker, P. Guillemin, and S. Woelfflé,Vision and Challenges for Realising the Internet of Things .Pub.Office EU, 2010 [Online]. Available: http://www.internet-of-things-research.eu/pdf/IoT_Clusterbook_March_2010.pdf

[6] V. Ovidiu and F. Peter, Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems , 2013.

[7] IT Union, "ITU Internet report 2005: The Internet of Things," 2005.

[8] Y. Zhang, M. Chen, S. Mao, L. Hu, and V. Leung, "CAP: crowd activity prediction based on big data analysis," IEEE Network, vol. 28, no. 4, pp. 52–57, 2014.

[9] J. Wan, D. Zhang, Y. Sun, K. Lin, C. Zou, and H. Cai, "VCMIA:a novel architecture for integrating vehicular cyber-physical systems and mobile cloud computing," Mobile Networks and Applications, vol. 19, no. 2, pp. 153–160, 2014.

[10] A´ lvaro Villalbaa, Juan Luis Pe´reza, David Carreraa, Carlos Pedrinacib, Luca Panzierab servIoTicy and Serve: a Scalable Platform for Mining the IoT Procedia Computer Science Volume 52, 2015, Pages 1022-1027

[11] Juan I. Guerrero, Antonio García, Enrique Personal, Joaquín Luque, Carlos León "Heterogeneous data source integration for smart grid ecosystems based on metadata mining" Elsevier Volume 79, 15 August 2017, Pages 254–268.

[12] Introduction to Data Mining and its Applications by S. Sumathi, S.N. Sivanandam ol. 20. Wojciech Penczek, Agata Pórola Advances in Verification of Time Petri Nets and Timed Automata, 2006 ISBN 3-540-32869-6

[13] N. Ali and M. Abu-Elkheir, "Data management for the internet ofthings: Green directions," inProc. IEEE Globecom Workshops, 2012,pp. 386–390.

[14] 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

[15] R. G. Baraniuk, "More is less: Signal processing and the data deluge,"Science, vol. 331, no. 6018, pp. 717–719, 2011.

[16] D. Reed, D. Gannon, and J. Larus, "Imagining the future: Thoughts on computing,"Computer, vol. 45, no. 1, pp. 25–30, 2012.

[17] M. Hilbert and P . L´opez, "The world's technological capacity to store, communicate, and computer information,"Science, vol. 332, no. 6025,pp. 60–65, 2011.

[18] S. Madden, "From databases to big data," IEEE Internet Computing,vol. 16, no. 3, pp. 4–6, 2012.

[19] Y. K. Chen, "Challenges and opportunities of internet of things," in Proc. Asia and South Pacific Design Automation Conference, 2012,pp. 383–388.

[20] D. Bandyopadhyay and J. Sen, "Internet of things: Applications and challenges in technology and standardization," Wireless Personal Communications, vol. 58, no. 1, pp. 49–69, 2011.

[21] V. Mayer-Schonberger and K. Cukier, Big Data: A Revolution ThatWill Transform How We Live, Work, and Think. Boston: Houghton Mifflin Harcourt, 2013.

[22] Libelium Comunicaciones Distribuidas S.L., "50 sensor applications for a smarter world," 2012, available at http://www.libelium.com/top 50 iot sensor applications ranking.

[23] R. Macmanus, "Top 10 internet of thingsproducts of 2009," 2009, available at http://www.readwriteweb.com/archives/top 10 internet of things products of 2009.php.

[24] U. M. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From data mining to knowledge discovery in databases," AI Magazine, vol. 17, no. 3, pp.37–54, 1996.

[25] Shen Bin, Liu Yuan, Wang Xiaoyi "Research on Data Mining Models for the Internet of Things" 978-1-4244-5555-3/10/$26.00 ©2010 IEEE

[26] Data Mining for the Internet of Things: Literature Review and Challenges, ACM digital library Volume 2015, January 2015 Article No. 12

[27] Ming –Syan chang, Jiawel Han, Philip S. yu "data Mining: an overview from database perspective "IEEE Transaction on knowledge and data engineering vol 8,no 6 December 1995.

[28] 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 77

[29] Jie Lin_, Wei Yuy, Nan Zhangz, Xinyu Yang_, Hanlin Zhangx, and Wei Zhao "A Survey on Internet of Things: Architecture, Enabling Technologies, Security and Privacy, andApplications" 2327-4662 (c) 2016 IEEE internet of things journals

[30] Rolf H. Weber "Internet of Things – New security and privacy challenges"

[31] Elsevier science direct available at www.sciencedirect.com

[32] Internet of Things and Data Analytics Handbook edited by Hwaiyu Geng

[33] Ovidiu Vermesan & Peter Friess "Internet of Thjings –Converging Technologies for smart Environment and integrated Ecosystem" , River Publication ,ISBN 978-87-93379-20-6

[34]]C. Hsinchun, W. Chung, J. J. Xu, G. Wang, Y. Qin, M. Chau,"Crime data mining: a general framework and some examples", Computer, 37(4), pp.50-56, 2004.

[35] Mohammad Abdur Razzaque, Member, IEEE, Marija Milojevic-Jevric, Andrei Palade, and Siobhán Clarke "Middleware for Internet of Things: A Survey" IEEE INTERNET OF THINGS JOURNAL, VOL. 3, NO. 1, FEBRUARY 2016

[36] Anne H. H. Ngu, Mario Gutierrez, Vangelis Metsis, Surya Nepal, and Quan Z. Sheng, Member, IEEE

[37] "IoT Middleware: A Survey on Issues and Enabling Technologies" IEEE INTERNET OF THINGS JOURNAL, VOL. X, NO. X, MARCH 2016 1

[38] www.prezi.com

[39] Ivan Kholod ,Mikhail Kuprianov ,Ilya Petukhov "Distributed data mining based on actors for Internet of Things" IEEE, Embedded Computing (MECO), 2016 5th Mediterranean Conference on 01 August 2016 DOI: 10.1109/MECO.2016.7525698

[40] Farheen Fatima1, BatulNaeem Husain2, Mohammed Azharuddin3, MohammedAbdul Mabood "Internet of things:A Survey on Architecture, Applications, Security, Enabling Technologies, Advantages & Disadvantages "International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12, December 2015Copyright to IJARCCE DOI 10.17148/IJARCCE.2015.412117

[41] O. Kurasova, V. Marcinkevičius, V. Medvedev, A.

[42] Rapečka, P. Stefanovič "Data Mining in Big Data Era: Methods and Technologies" 6th International Workshop DATA ANALYSIS METHODS FOR SOFTWARE SYSTEMS https://www.mii.lt/datamss/files/liks_mii_drusk_2014_abstract_last_print_.pdf.