Integration of Data Mining and IoT

Deepali Joshi

[1]*(Computer Engineering, Thakur College Of Engineering & Technology/ Mumbai University, India)*
Corresponding Author: Deepali Joshi

Abstract: *In today's world number of devices are used in order to complete the work. These devices generate a lot of data and also need to communicate with other devices. It is necessary to analyze the data generated by these machines. When devices want to connect it is necessary to have an internet connection. IoT is basically a network which consists of devices, vehicles, home appliances, software's, sensors, actuators, and they are connected to collect and exchange the data. In order to achieve this data mining and IoT can be utilized. The collaboration of IOT and Data Mining will make various tasks simpler. Merger of data mining and IoT is discussed in this paper. This paper represents Data mining challenges, and Data mining issues with IOT.*

I. Introduction

IoT is the process of connecting each and every thing through the Internet. Recently many new applications and services are introduced. IOT will help to connect number of nodes of various objects with the web servers and supercomputers. New technologies can be integrated very easily with IOT. Nowadays everyone has devices such as mobile phones, with the help of these devices, there is no limitation that corrupts the connection between people. Researchers working in various fields such as academics, government have interest in modifying the internet by designing various systems such as smart home, smart pen, intelligent transportation, global supply chain, healthcare.

II. Data from IoT

IOT creates data by gathering information from various devices. Sensors are used to gather data and should be able to collect useful information. It is essential to handle large amount of data obtained from devices and to find useful information. Knowledge Discovery in Databases (KDD) is used to extract data and has following steps: data gathering, data cleaning, data integration, data transformation, data mining and evaluation or decision making. Data extraction is done at Data mining step.

III. Data Mining from IoT

IOT It is essential to understand relationship between data mining and IOT. The very first step is to capture the data from devices such as Sensor networks, Actuators, WSN (Wireless Sensor Network), WSAN (Wireless Sensor and Actuator Network), RFID (Radio Frequency Identification)Tags, Camera, GPS etc. These devices generate large amount of data and it will be in different formats. The devices that generate and capture data are as follows:

1. Sensor Networks

A sensor network is a group of small devices and which will monitor and store the conditions and data. Sensor networks are used widely in industry, health industry, manufacturing, scientific and engineering applications, education industry, home appliances etc. it comprises of wireless sensor networks.

2. WSN (Wireless Sensor Networks)

WSN is a network of various autonomous nodes which covers a particular region, gathers data and provides the information. There are number of wireless sensor nodes which have some computational power, some memory, limited bandwidth and sensing capability. They capture and store the data of the environmental conditions and send the data to a sink or a base station which further processes and analyzes the data. The sensor nodes can communicate with each other using radio signals.

3. RFID

RFID means Radio Frequency Identification. It used to identify objects and people using radio technology of short range. These objects and people also can communicate with each other digitally. There are 2 entities here: Reader which is stationary and Tag which is movable. This tag on any object is being read by the

reader to identify it. RFID consists of: transceiver (transmitter/receiver) and antenna are usually combined as an RFID reader. A transponder (transmitter/responder) and antenna are combined to make an RFID tag [8]. Tags are scanned and information is forwarded to backend. The backend consists of database, application interface. When information is received, backend adds it to database.

Example of RFID systems and implementation is supermarket chains. All the products is the shopping bag of the customer may be identified by RFID reader as all the products have RFID tags on them. After identifying the products, the information is send to the backend which further provides the additional information such as prices of the products, discounts on the qualifying products and for the qualifying customers, etc. Also the backend reduces the number of products in the database and notifies the concerned authority if the stock of the products have to be updated.In many applications, barcodes are often used but the difference between bar codes and RFID systems are RFID systems do not need a line of sight, RFID readers may scan many products at one time and they have more storage than barcode systems.

IV. How do Data Mining Processes Work
IoT data mining processes are divided into different stages, as follows:
1. Data is integrated according to different data sources.
2. Data is cleaned, so it can be easily extracted and processed.
3. Some parts of data are extracted and prepared for future processing.
4. Sophisticated algorithms are used to identify patterns.
5. Data is restructured and presented to the users in a coherent way.

V. Challenges in Data Mining with IoT
1. A major challenge is to upgrade a crime detecting application which includes advanced features that prevent crime.
2. One challenge is to extract large data available in large data storage and to detect any noise or unreliable data in that large dataset.
3. Mining uncertain and incomplete data is also a big data challenge. The algorithm modification is also very difficult, providing security solutions for sharing data is not so easy [11].
4. Its is difficult to handle and analyse large data for data mining purpose.
5. Building an intelligent system with the help of simple algorithm is not possible. For preparing an intelligent system, a number of algorithms must be fused to a single algorithm .
6. Internet connection may arise as a challenge when used in rapid speed device.
7. It is a challenge to build up an intelligent industrial IOT device, which includes smart city, green energy generation etc.
8. Major challenges in security are widespread data collection, unexpected use of   consumer data.
9. Sharing of standards and infrastructure gives rise to a security issue.

VI. Major Issues in Data Mining with IoT
1. Parallel programming needs to be designed in such a way that every algorithm can be applied to it].
2. The frame work that considers security, privacy, data sharing, the growth of   data size etc is to be designed.
3. Infrastructure Perspective: IOT gives low computation and high throughput,   but mining algorithm is designed for small size & low power consumption device. This creates an infrastructure issue.
4. Data Perspective: Gathering data from different sources creates redundant data. The user needs to filter the redundant data for better system performance. Moreover, data generate from different from sources may become an obstacle.
5. Algorithm Perspective: Some example of IOT needs adding classifiers dynamically and some needs adding classifier statically. So, we need some mining technologies to be fused that can classify the classifiers in a common way.
6. Privacy and Security: Privacy and Security remains an issue because every algorithm and technologies are not able to outperform privacy and security issue. For example, it is easy for companies to collect different customer data from various devices or sources and use data mining techniques to find   the information that helps in increasing sales volume but the issue is that many customers wouldn't like to disclose their privacy and security, such as   retail, shopping behavior.
7. Massive Scaling: How to name, identify, authenticate, maintain, protect and use a large number of data is an open issue.
8. Architecture and dependency: It is difficult to construct an architecture that connects a large number of things to the internet. Many things are dependent on each other, so removing any of them may generate an error[18].

9. Robustness: In robustness, devices location have to be known if clock drifts, the location of the device may not be accurate.

VII. Conclusion

The usage of devices has increased and due to this a lot of data is generated by these devices. In order to deal with devices and data generated by them, data mining techniques are essential. The fusion of IOT with data mining helps to discover useful information from large amount of data. Various algorithms are available that can be applied on large datasets. The techniques of Data mining are clustering, classification, pattern mining. This paper focuses on the challenges and issues of data mining. Since the IoT is in the initial stage, focus is on developing efficient mechanism and algorithms.

References

[1]. Feng Chen, Pan Deng, Jiafu Wan, Daqiang Zhang, Athanasios V. Vasilakos, and XiaohuiRong, "Data Mining for the Internet of Things: Literature Review and Challenges", Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, 2015.

[2]. Chun-Wei 12.

[3]. Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang, "Data Mining for Internet of Things: A Survey", IEEE Communications Surveys and Tutorials, 2014.

[4]. Shweta Bhatia, Sweety Patel, "Analysis on different Data mining Techniques and algorithms used in IOT", International Journal of Engineering Research and Applications, November 2015.

[5]. John A. Stankovic, "Research Directions for the Internet of Things", IEEE, 2014.

[6]. Shen Bin, Liu Yuan, Wang Xiaoyi, "Research on Data Mining Models for the Internet of Things", IEEE, 2010.

[7]. Yan Chen, Ai-xia Han Cai-hua Zhang, "Research on Data Mining Model in the Internet of Things", International Conference on Automation, Mechanical Control and Computational Engineering, 2015.

[8]. H. Chen, W. Chung, J. J. Xu, G.Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," Computer, vol. 37, no. 4, pp. 50–56, 2004.8] X. Wu and S. Zhang, "Synthesizing high-frequency rules from different data sources," IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 2, pp. 353–367, 2003.

[9]. X. Wu and S. Zhang, "Synthesizing high-frequency rules from different data sources." IEEE Transactions on Knowledge and Data Engineering, vol. 15, no.2. pp. 353-367, 2003.

[10]. T. Keller, "Mining the internet of things: Detection of false-positive RFID tag reads using low-level reader data," Ph.D. dissertation, The University of St. Gallen, Germany, 2011.

[11]. A. de Saint-Exupery, "Internet of things strategic research roadmap," European Research Cluster on the Internet of Things, Tech. Rep., 2009.

[12]. R. A. Baeza-Yates and B. A. Ribeiro-Neto, Modern Information Retrieval. ACM Press Addison-Wesley, 1999.

[13]. T. He, J. Stankovic, C. Lu and T. Abdelzaher, "A Spatiotemporal Communication Protocol for Wireless Sensor Networks", IEEE Transactions on Parallel and Distributed Systems, Vol. 16, No. 10, Oct. 2005.

[14]. B. Liu,"Web Data Mining: Exploring Hyperlinks, contents, and Usage Data". Berlin, Heidelberg: Springer-Verlag, 2007.

[15]. K.Ashton, "That 'Internet of Things' Things,"2009, RFID Journal, Available at http://www.rfidjournal.com/article/print/4986.

[16]. Auto-ID Labs, Masschusetts Institute of Technology, 2012, Available at http://www.autoidlabs.org/.

[17]. L.Atzori, A. Iera, and G. Morabito,"The internet of things: A survey," Computer Networks, vol.54, no.15, pp. 2787-2805, 2010.

[18]. D. Miorandi, S.Sicari, F. DePellegrini, and I.Chlamtac,"Internet of things: Vision, applications and research challenges", Ad Hoc Networks, vol.10,no.7, pp.1497- 1516,2012

[19]. 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.

[20]. M.C. Domingo, "An overview of the internet for things for people with disabilities," Journal of Network and Computer Applications, vol.35, no,2, pp.584- 596, 2012.