International Journal of Innovative Research in Engineering
Volume 3, Issue 5 (September-October 2022), PP: 143-147.
www.theijire.com
ISSN No: 2582-8746

Data Mining In IoT

V. Iswarya[1], P. Jaya Shree[2], N. Pooja[3], K. Shreenidhi[4], R. Selvapriya[5], A. Puviyarasan[6]
[1,2,3,4,5,6] Computer Science, Sri Gvg Visalakshi College For Women, BharathiarUniversity, Tamilnadu, India.

How to cite this paper: V. Iswarya[1], P. Jaya Shree[2], N. Pooja[3], K. Shreenidhi[4], R. Selvapriya[5], A. Puviyarasan[6], "Data Mining In Iot", IJIRE-V3I05-143-147.

Abstract: *This paper provides an overview on Data Mining (DM) technologies for the internet of things (IoT). Data Mining is one of the most valuable Technologies enable to identify Unknown patterns and make internet of things smarter. The huge network of devices generates a new type of data known as the IoT big data. The biggest challenge in today's data mining world comes with several issues like data storage, management, privacy, security, and processing limitations such as real time streaming data. DM techniques applicable to different IoT data. Some data specific features were analyzed, and algorithms for knowledge discovery in IoT data were considered. Data mining algorithms can be applied to IOT to extract hidden information from data. Various data mining models have been proposed for Internet of Things. This paper provides a systematic and detailed review of various data mining Techniques employed in the large and small scale IoT and also discuss about the main datamining techniques, big data mining issues in IoT, applications of IoT in data mining. And also cover the issues of how the applications of data mining in IoT enables the useful accumulation of tremendous amounts of data generated from heterogeneous environments and transformed into valuable insights that can be used to the great benefit of the organization.*

Key Word: *Data mining, Internet of Things, Data preprocessing, Data cleaning, Data collection, PredictiveAnalysis in IoT.*

I. INTRODUCTION

In order to make IoT smarter, lots of analysis technologies are introduced into IoT; one of the most valuable technologies is data mining. IoT has been playing an essential role ever since it appeared which covers from traditional equipment to general household objects. Application of data mining in IoT offers comprehensive information about the connected assets and their data interchange- the baseline (normal communication between the devices); data mining is used when you need to customize your baseline and use the data insights for your specific business needs. The growth of IoT is propelled by the increasing availability of micro-computers, smart phones, wearable and domoticgadgets. Such smart devices are expected toreach 30 billion by 2020, giving rise to enormous amounts of valuable data, services, and business opportunities the business models and applications that will be powered by 5G represent a significant opportunity for operators. Ericsson expects that in 2026, there will be a USD 582 billion global markets as telecom operators leverage 5G technology for industry digitalization. IoT applications generate more than 2.5 quintillion data bytes daily. To convert this data into knowledge, data mining systems are increasingly in demand. The expectations of IoT and its relevant products in this new era are quite high. A new data mining approach is edge computing that works on wireless, battery- power, smart sensing devices that sit at the edge points of the IoT. Edge computing/ edge mining can minimize the number of packets transfer in order to reduce data and energy requirements. The instability of the network connection and availability of the thing due to the unpredictable mobility of devices, different battery discharge rates, equipment failures and lack of a priori knowledge of the hardware and software characteristics of devices. Data mining has a vast application in big data to predict and characterize data. The Internet of Things (IoT) has become an important research domain as mature appliances, systems, infrastructures, and their potential in recent years.

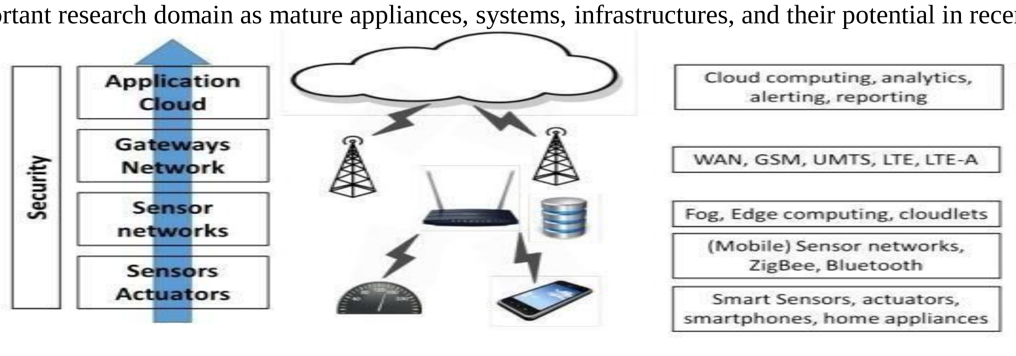

Figure 1: IoT layers

Our main contribution in this paper is that we targeted on data specific features and selected some well-known algorithms best suited for knowledge discovery in different IoT applications.

II. DATA MINING TECHNOLOGIES FOR IOT

Extracting useful information from a complex sensing environment at different spatial and temporal resolutions is a challenging research problem in artificial intelligence. To detect useful patterns in IoT data, the data needs to be analyzed with suitable DM techniques. DM is highly domain specific. Whenever the IoT platform has to predict optimal traffic routers or detect a machine that is about to fail and needs maintenance, different methods might apply. For instance, a predictive maintenance application that needs to detect machine failure before it occurs so it can be replaced before production is interrupted, collects and analyses sensor data from machines such as temperature, torsion, or attrition. A home security application might use movement detectors and camera data to detect possible intruders. Both systems use different types of data and make predictions about different things. Also, IoT has some characteristics that influence the methods with which IoT data is analyzed:

- New devices can be added ad hoc to an IoT solution. This means that new data sources need to be analyzed, possibly in new data formats. for instance, an e Health application can measure blood pressure and glucose levels. A patient using the solution has a new fitness tracker that can add fitness data to the e Health solution.
- Devices might stop sending data. For instance, a car is driving into a tunnel and loses the GPSsolution.
- A sensor might stop sending data at all because the battery is empty or the wireless communication isinterrupted.
- A sensor or actuator can be part of several applications. For instance, a movement detector can be used toopen an automatic door and to detect unauthorized intruders, if the detector fails, several applications are affected. An IoT application might have to show a different behavior in different situations. For instance, a home security system has to be able to differentiate between day and night, since during the day a greater number of human activities is detected, at night hardly any.
- The ramifications are derivations in normal data traffic that might be interpreted as anomaly and the IoT solution might issue a false alert event. Dm techniques for IoT need to be able to adapt to dynamic environments or changed data streams to avoid redesign of the DM rules each time a sensor is added or removed. Machine learning (ML) techniques are well suited to handle the fuzziness in data streams and can adapt quickly when the environment changes. ML is a branch of artificial intelligence (AI)and aims to imitate human learning on computers without the need to be explicitly programmed.

ML techniques learn the DM rules from historic data so there, there is no need for a developer to program them manually. ML techniques have several characteristics that make them favorable for IoT DM:

- ML methods can continue to learn new rules, for instance if a new smart device is added.
- Many ML schemes calculate probabilities, which makes them robust against small changes in thedata flow. For instance, when a device stops sending data and there are still others sending values,the probability changes only slightly and no false positive is issued.

Data mining goes through several steps. They are divided into a data conditioning or data preprocessing phase and a predictive analysis phase. In the data conditioning phase, data is collected and preprocessed. Not all data is useful for a specific DM task so

Figure 2: Data Mining Processes

Observation points are an important preprocessing step. other preprocessing step include data transmission to have consistent data format ,data formats ,data duplication and outlier removal .In the prective analysis phase, suitable DM methods have to be selected and trained. Depending on the problem, the data has to be correlated since data from only one data source might not be enough to make meaningful predictions.

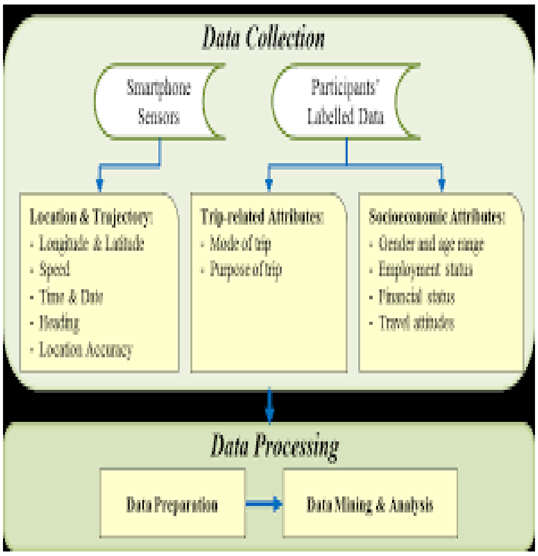

Figure 3: data collection and data processing.

III.DATA COLLECTION

Data collection in IoT happens at the device or "thing" level. Smart sensors, smart phones or tablets measure different values from the environment and typically transmit them to an IoT cloud platform for analysis and storage. Some devices produce potentially a lot of data and transmitting all data to the IoT cloud might not be practicable. To save bandwidth and computing power, often the data is stored and preprocessed in the fog by edge computers. Edge computing often **DATA MINING.**

Data mining is the process of sorting through large data sets to identify patterns and relationships that can help solve business problems through data analysis. Also (pre-) analyses data and only some observation points are transmitted to the cloud.

Figure 4: data collection

IV.DATA PRE-PROCESSING

Real life data is seldom in a format suitable for DM algorithms and is often poor in quality. Data cleaning is thus a crucial step to get good results. Different sensors collect data in different formats. Data transformation is an important step to harmonize data. Relevance filtering is another important preprocessing step to get good performance of IoT applications. For instance, one application might only require the source and destination coordinates of an itinerary, for another application the whole route might be relevant. Data de duplication, outlier removal, entity resolution and feature selection are some important preprocessing steps. Feature selection are some important preprocessing steps. Feature selection will mean selecting the observation points that are used as input for the DM algorithms.

Data mining techniques and tools enable enterprises to predict future trends and make more-informed business decisions. There are many DM techniques. Machines learning (ML) techniques are adopted when the rules are getting too complex or if there are too many rules to be programmed by a developer.ML imitate human learning. Whereas humans learn from experience, algorithms learn the rules from historical data. ML techniques learn from past, historical data, to make predictions about future events. For instance, a predictive maintenance application uses historic sensor data that collects information about the state of a smart building to learn rules to predict if the air conditioning system or elevators are going to fail. ML is divided into semi- supervised and unsupervised learning. Supervised methods are used for classification and regression. They require labeled data for training. Typical supervised learners include Bayesian models, decision tree induction, support vector machines (SVM) and Artificial Neural Networks (ANN). Unsupervised methods are adopted if no labeled data is present and semi- supervised methods are used when a small amount of labeled data and a large amount of unlabeled data is available. The advantages of data mining include making better decisions, having a competitive advantage, and finding major problems.

V.PREDICTIVE ANALYSIS

Predictive analysis refers to the use of both new and historical data, statistical algorithms, and machine learning techniques to forecast future activity, patterns, and trends. The primary objective is to go beyond knowing what has happened to assess better what will happen in the future. As the name suggests, „predictive" means to predict something, so predictive analytics is the analysis done to predict the future event using the previous data. It is the process of extracting information from existing sets of data to find useful information, trends and forecast future events. Predictive analytics does not tell the exact thing that will happen in the future. It predicts what might happen in the future.
These are the given most significant business benefits of predictive analytics.

1. Predictive analytics increase production efficiency.
2. It minimizes business risks.
3. It helps in decision-making purposes in any business organization.
4. It drives a competitive environment.

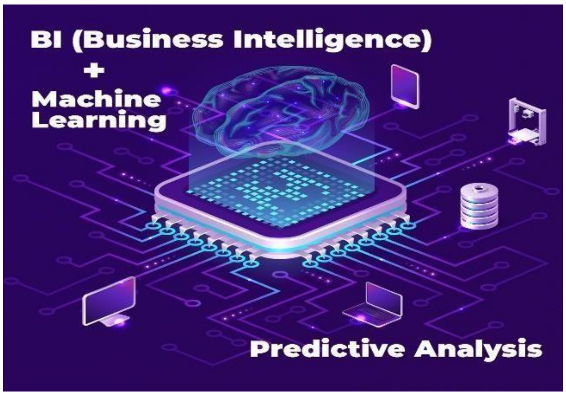

Figure 5: predictive analysis

Based on the information gathered from predictive analytics, many companies have increased their turnover, achieved goals, and increased revenues by applying these strategies.

1. Synchronizing supply with demand.
2. Fraud detection.
3. Creating lasting inventories.
4. Customer satisfaction.
5. Setting prices appropriately In order to increase profit.

VI.DATA MINING APPLICATIONS

Data Mining in e-Commerce

Data mining enables the businesses to understand the patterns hidden inside past purchase transactions, thus helping in planning and launching new marketing campaigns in prompt and cost-effective way.

Data Mining in Industry

Data mining can highly benefit industries such as retail, banking, and telecommunications; classification and clustering can be applied to this area. One of the key success factors of insurance organizations and banks is the assessment of borrowers, credit worthiness in advance during the credit evaluation process.

Data Mining in Healthcare

In healthcare, data mining is becoming increasingly popular, if not increasingly essential. Heterogeneous medical data have been generated in various health care organizations, including payers, medicine providers, pharmaceuticals information, prescription information, doctor's notes, or clinical records produced day by day.

Data Mining in City Governance

In public service area, data mining can be used to discover public needs and improve service performance, decision making with automated systems to decrease risks, classification, clustering, and time series analysis which can be developed to solve this area problem. A major challenge for the government and law enforcement is how to quickly analyze the growing volumes of crime data. Researchers introduce spatial data mining technique to find out the association rules between the crime hot spots and spatial landscape.

VII.CHALLENGES

One of the biggest challenges in IoT is the lack of standards which makes interoperability of different devices and connecting them to the internet difficult. Standardization efforts and standard protocols such as Message Queuing Telemetry Transport (MQTT) and Advanced Message Queuing Protocol (AMQP), which are light-weight Message oriented middleware, have emerged. Nevertheless more data transformation is needed.

Figure 6: prediction in data mining

Many smart devices have limited resources, limited bandwidth and battery life. Also, mobile networkcoverage varies in different places which can be problematic especially for e Health IoT applications. Security remains a major concern in IoT since many devices were not designed with security in mind and they remain vulnerable to cyber-attacks. Due to their limited nature, securing them with encryption and intrusion prevention mechanism remains challenging. Privacy is a major concern, especially in places like smart homes where a great deal of personal data is collected. Privacy preserving DM techniques have been proposed, but due to the restrictions of the devices and all the other challenges in IoT DM they are often not applied. Privacy and AI are complex topics and more research in privacy engineering is high desirable.

VIII.FUTURE DEVELOPMENT

Technology becomes an integral part of our environment and the volumes and heterogeneity of the data that need to be processed will never increase. To be useful, IoT data often needs to be analyzed in real- time, e. g. to respond to traffic jams or to optimize energy consumption. The demand for more and faster resources will increase and more solutions will be operated in the cloud. Cloud solutions provide the scalability that many IoT applications require. An emerging area in ML is deep learning. Contrary to the learners described above, deep learners (DL) have multiple layers of abstraction to interpret data. As the signal processes through the layers, the data is represented at a higher abstraction level. This makes DL more robust if a data stream shows deviations from normal streams, which might be interpreted as an anomaly by traditional ML techniques, also called shallow learners and also some DL can do feature extraction automatically, there is no need for feature engineering. This will lead to IoT solutions that can prioritize and thus learn what is relevant indata and what is not. This allows DL enabled IoT solutions to learn from new situations and react accordingly. For instance, in a traffic situation an IoT enabled car can decide if it should evade or break. DL have been particularly well performing for multimedia mining and natural language processing (NLP) and we will see some more IoT expand into new areas. IoT is used to form smart homes and smart cities.Ultimately this will leadto the smart planet that will become smarter, more efficient, greener, more economical, and will self- regulate the energy consumption, waste disposal, improve security and overall quality of life.

IX.CONCLUSION

The internet of things concept arises from the need to manage, automate, and explore all devices, instruments, and sensors in the world. In order to make wise decisions both for people and for the things in IoT, data mining technologies are integrated with IoT technologies for decision making support and system optimization. Data mining involves discovering novel, interesting, and potentially useful patterns from data and applying algorithms to the extraction of hidden information.

References

1. *Agrawal. R & Srikant. R 2000,"privacy- presenting data mining", ACM Sigmod Record, Vol. 29, no. 2,pp. 439-50.*
2. *Arias. M, Arratia. A & Xuriguera. R 2014," forecasting with twitter data", ACMTrans. Intell. Syst. Technol.,vol. 5, no.1, pp. 1-24.*
3. *Cubo. J, Nieto, A & Pimente. E 2014, "A Cloud-Based Internet of Things Platform for Ambient Assisted Living",Sensors (14248220), vol. 14, no. 8, pp. 14070-105. [4] Dohr. A, Modre-Opsrian, R, Drobics.M,Hayn.D & Schreier, G 2010, "The Internet of Things for Ambient Assisted Living", in Information Technology: New Generations (ITNG), 2010 Seventh International Conference on: proceedings of the Information Technology: New Generations (ITNG), 2010 Seventh International Conference onpp. 804-9.*
4. *Zhihao, X & Yongfeng, Z 2010, "Internet of Things and its future", HuaweiTechnologies Communicate BeyondTechnology, pp. 23-6. [17] Zhong, N, Ma, J, Huang, R, Liu, J, Yao, Y, Zhang, Y & Chen, J 2016, "Research challenges and perspectives on Wisdom Web of Things (W2T)", in Wisdom Web of Things, Springer, pp. 3-26.*
5. *Soyturk, M, Muhammad, KN, Avcil, MN, Kantarci, B & Matthews, J 2016,"Chapter 8 - From vehicular networks to vehicular clouds in smart cities A2 - Obaidat, Mohammad S", in P Nicopolitidis(ed.), Smart Cities and Homes, Morgan Kaufmann, Boston, pp. 149-71.*
6. *Islam, SR, Kwak, D, Kabir, MH, Hossain, M & Kwak, K-S 2015, "The internet of things for health care: acomprehensive survey", IEEE Access, vol.3, pp. 678-708*