Role of Data Mining in Developing a Smart Iot and its Challenges

Nandhini.S, Jeen Marseline .K.S, Deepa.B

Abstract: ***Internet of Things (IoT) is a rapidly growing technology on which enterprises are building their future. IoT will change everything into "smart" ranging from smart home to smart government. To make them smart the various IoT devices are used which act as sensors and generate vast amount of data. The right IoT technology will aggregate, process and interpret the data generated by smart devices. The process of information extraction from a very large database is difficult. To make the data useful, which is collected from, a variety of devices, it is appropriate to use data mining techniques. Using various data mining techniques, we can extract data from various sources in an effective manner. This paper gives an overview of different data mining techniques, which can make IoT devices smarter and also addresses some challenges faced in applying these techniques.***

Index Terms: IoT, Data Mining, Classification, Logistics Regression, Naïve Bayes, K- Nearest Neighbors

I. INTRODUCTION

IoT is a set of interrelated devices, which have unique identifiers, have sensors in them and are capable of computing. These devices can be connected to internet via wireless and wired internet connections. IoT is capable of connecting both living and lifeless things. IoT devices can be categorized into wearable and embedded. The data collected from IoT can be processed and used for various purposes [1].The information of the objects in IoT are stored in a repository where the hidden data need to be extracted in a smarter way, thus the data mining techniques are used. The extracted data can be classified and clustered.

In the day to day life huge amount of data is being stored in the database. These data contain knowledge about several aspects of business decision support. Data mining is the process of discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems("Data Mining",n.d.). The Machine Learning technology uses number of techniques on different models of information being stored in the database. Since the data sets are large, the efficiency and scalability of the algorithms are important. From the different techniques of Data Mining algorithms, Classification and Clustering algorithms are mainly focused in this paper relating to IoT.

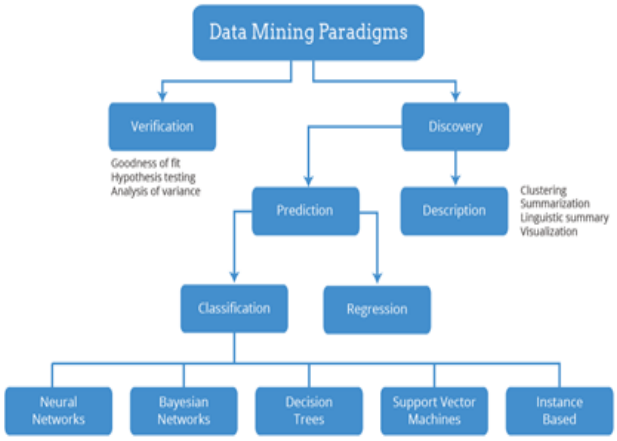

Fig I: Data Mining Overview

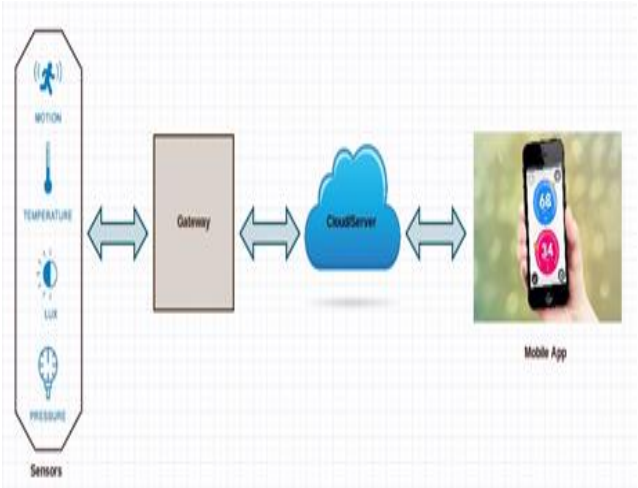

Fig II: IoT Architecture

Revised Manuscript Received on December 22, 2018.

Nandhini.S Assistant Professor,Dept of Computer Technology Sri Krishna Arts and Science College,Coimbatore,India nandhinis@skasc.ac.in

Jeen Marseline .K.S Assistant Professor,Dept of Information Technology,Sri Krishna Arts and Science College Coimbatore,India,Jeenmarselineks@skasc.ac.in

Deepa.B Assistant Professor,Dept of Information Technology Sri Krishna Arts and Science College,Coimbatore,India

II.LITRATURE REVIEW

Since IoT provides heterogeneous and sparse data, an effective tool called data fusion can be used to manage the data with efficiency. Jin Zhou et al. [5] proposed an algorithm which uses the technique of partitioning. In this work a higher dimension data set is converted into smaller data subsets for easy processing. Then based on the partitioning core attribute sets are obtained. The results are obtained by two different methods namely attribute reduction and rule extraction.

In order to enhance the performance of wireless IoT network, the heterogeneous data which is gathered from huge set of data must be efficiently analyzed. The two domains namely edge computing and cloud computing can be combined together for doing data analytics in real time. Shree Krishna Sharma and Xianbin Wang [6] proposed a framework which integrates the both "cloud computing and edge computing" platforms to get the advantages from these two paradigms. They also proposed extended research work in this area.

When analytic power is added to IoT, it becomes a valuable resource. Mohsen Marjani et al. [7] proposed a new architecture which can be applied for big data analytics. A comparative study of various analytic types and their levels are done in their work. They have also presented various use cases for IoT data analytics.

BiljanaL et al. [8] proposed a holistic framework to integrate different smart home objects as an effective, smart IoT based solution. They have also identified a "smart home management model" for the identified framework. They have considered energy management and architectural challenges of smart home in order to arrive at the holistic approach.

The role of IoT in smart city is very significant. To develop smart cities, Advanced Metering Infrastructures can be integrated with IoT for deployment smart cities. Lloret, J. et al. [9] proposed the architecture and communication protocols for the deployment of smart meters which will provide utility information to the people efficiently

III. CLASSIFICATION ALGORITHMS

Since IoT produces heterogeneous data we need to categorize them to make data more useful. The process of classification can be performed either on structured or unstructured data. Classification is the process of assigning items in a collection to a number of classes. The main role of classification is to identify the appropriate class in which the data should be stored.

The following are the few terminologies used in classification function of machine learning:

Classifier: It is the algorithm that makes use of some training data to understand how input data relate to a class.

Classification model: It will give the conclusion for the input values.

A. Logistic regression

It is a kind of machine learning algorithm used for classification technique. It is used to determine in what class a new data should fall into. It is used to classify the problems like email (spam/not spam), online transaction (fraudulent/not). This algorithm will have a variable, which can assign only two values either positive class or negative class. Logistic regression has a number of applications in the industry. In the field of IoT, these techniques can be applied for predicting equipment failure based on the details gathered

B. Naïve bayes algorithm

This algorithm is based on Bayes theorem. It is based on probabilistic model.It includes independent assumptions. It is a classification algorithm. This model is easy to build. It is very useful in huge data set. In IoT environment, we have very huge data set which is of heterogeneous nature. In order to classify those data effectively and efficiently we can use this algorithm.

ADVANTAGES:

- Very simple, easy to implement and fast.
- Need less training data.
- It is used for different kinds of classification problems like binary and multiclass classification.
- It handles continuous and discrete data.
- Not sensitive to irrelevant features.

DISADVANTAGES:

- Given the output class, the two features are independent. So the result will be bad.
- Large amount of data scarcity
- Due to continuous features some information can be lost in large amount.

C. K-Nearest Neighbors

It is the most essential algorithm in machine learning and IoT. It is non-parametric. That is, it does not make assumptions with the distributed data. KNN is a very simple classification technique. KNN is efficient even when there is little or no prior knowledge about the distribution of the data is available [3]. This technique assigns the property value for the object as the average of the values of its K nearest neighbors. This algorithm has a variety of applications in the field of IoT like intrusion detection and activity recognition [5].

ADVANTAGES:

It is very useful when the training data is large.

It is more robust against noisy data.

DISADVANTAGES:

The value of the nearest neighbor must be known for this technique.

It includes a high cost of computation.

IV.CHALLENGES IN IOT

The data mining algorithms we have discussed above are useful in implementing a smart IoT system but still there are several challenges to be faced while implementing them.

The major challenge is the presence of large data set which makes the system very complex. Comparison between data sets becomes difficult since they come from different sources. Some applications require real time data processing and analysis which is very difficult for a large data set. In future these challenges will be overcome by appropriate new data mining

V . CONCLUSION

To make IoT applications much smarter, we need to combine the efficiency of data mining algorithms with vital data collected from IoT devices. They need to be analyzed with proper analytical tools to get the best usage of IoT technology. When these two paradigms are collaborated, we will be able to enjoy the benefits of smarter IoT applications

REFERENCES

1.Furqan Alama, Rashid Mehmoodb,∗, Iyad Katiba, Aiiad Albeshria

A Department of Computer Science, Faculty of Computing and Information Technology (FCIT), King Abdulaziz University, Jeddah, KSA International Workshop on Data Mining in IoT Systems (DaMIS 2016) Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)

2.Cao Dong1,2, Qiao Xiuquan2, Judith Gelernter1, Li Xiaofeng2, Meng Luoming2 1 School of Computer Science, Carnegie Mellon University, Pittsburgh, 15213, USA 2 State Key Lab of Networking and Switching Technology, Beijing University of Posts and

Telecommunications, Beijing, 100876, Mining Data Correlation from

3.Sadegh Bafandeh Imandoust And Mohammad Bolandraftar

Department of Economics,Payame Noor University, Tehran, Iran. Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background

4.Mr.Rakesh Kumar B 1, Maria Telma Dsouza 2 & Praveen U.C 3. A COMPARATIVE STUDY OF KNN AND SVM ALGORITHMS IN IOT

5.Jin Zhou, Liang Hu, Feng Wang, Huimin Lu, and Kuo Zhao An Efficient Multidimensional Fusion Algorithm for IoT Data Based

on Partitioning.

6.SHREE KRISHNA SHARMA, (Member, IEEE), AND XIANBIN WANG, (Fellow, IEEE) Department of Electrical and Computer Engineering, Western University, London, ON N6A 3K7, Canada

Corresponding author: S. K. Sharma (sshar323@uwo.ca) Live Data Analytics With Collaborative Edge and Cloud Processing in Wireless IoT Networks

7. MOHSEN MARJANI1, FARIZA NASARUDDIN2, ABDULLAH GANI1, (Senior Member, IEEE), AHMAD KARIM3, IBRAHIM ABAKER TARGIO HASHEM1, AISHA SIDDIQA1, AND IBRAR YAQOOB 1Department of Computer System and Technology, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia 2 Department of Information System, Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia 3 Department of Information Technology, Bahauddin Zakariya University, Multan, Punjab, Pakistan Corresponding authors: Fariza Nasaruddin (fariza@um.edu.my); Abdullah Gani (abdullah@um.edu.my) Big IoT Data Analytics: Architecture, Opportunities, and Open Research Challenges

8.Biljana L. Risteska Stojkoska ∗, Kire V. Trivodaliev Faculty of Computer Science and Engineering, University “Ss. Cyril and Methodius”, Skopje, Macedoni A review of Internet of Things for smart home: Challenges and solutions

9.Jaime Lloret, Jesus Tomas, Alejandro Canovas, Lorena Parra Instituto de Investigación para la Gestión Integrada de zonas Costeras.Universidad Politécnica de Valencia, Spain jlloret@dcom.upv.es,jtomas@dcom.upv.es, lcasol@posgrado.upv.es, loparbo@doctor.upv.es An Integrated IoT Architecture for Smart Metering

10.Machine learning for internet of things data analysis: a survey Mohammad Saeid Mahdavinejad a,b,*, Mohammadreza Rezvan a,b, Mohammadamin Barekatain c, Peyman Adibi a, Payam Barnaghi d, Amit P. Sheth b a University of Isfahan, Iran b Kno.e.sis - Wright State University, USA c Technische Universit€at München, Germany d University of Surrey, UK Digital Communications and Networks journal homepage: www.keaipublishing.com/en/journals/digital-communications-and-networks

11.Mr.Rakesh Kumar B1, Maria Telma Dsouza2 & Praveen U.C3A COMPARATIVE STUDY OF KNN AND SVM ALGORITHMS IN IOT

12.James Jin Kang, Sasan Adibi, Henry Larkin, Tom Luan School of Information Technology Deakin University Burwood Australia {jkang, sasan.adibi, henry.larkin, tom.luan}@deakin.edu.auPredictive Data Mining for Converged Internet of Things: A Mobile Health Perspective

13.George Suciu1 & Victor Suciu1 & Alexandru Martian1 & Razvan Craciunescu1,2 & Alexandru Vulpe1 & Ioana Marcu1 & Simona Halunga1 & Octavian Fratu1Big Data, Internet of Things and Cloud Convergence – An Architecture for Secure E-Health Applications