Research of various Data Mining Techniques for IOT Applications

Shaik Johny Basha, M.V.V.S. Subrahmanyam

Abstract— In recent years everything is connected and passing through the internet, but Internet of Things (IOT), which will change all aspects of our lives and future. While the things are connected to the internet, they will generate the huge amount of information which has to be processed. The information that gathered from various IoT devices has to be recognized and organized according to the environments of their type. To recognize and organize the data gathered from different things, the important task to be played is making things passing through different Data Mining Techniques (DMT). In this article, we mainly focus on analysis of various Data Mining Techniques over the data that has been generated by the IOT Devices which are connected over the internet using DBSCAN Technique. And also performed review over different Data Mining Techniques for Data Analysis.

Index Terms — IOT, Data Mining Techniques (DMT), K-Means Algorithm

I. INTRODUCTION

The IoT is a globally connected through the various devices in the world, enabling sophisticated services by intention and effective things based on existing technologies. It combines the different technology that combines the various networks, devices, mechanism and objects. Now a day's world is connected through the internet and uses same internet to connect the different devices, easily maintain and controlled all the operations. One of the most useful concepts is various Data Mining Techniques.

S.Haller et al. IoT definition given by "a world substantial object is seamlessly combined into the information network and the physical objects can become mostly active participants in business process". The huge amount of data being generated from various IoT devices is stored in different data sets. From the datasets discovering useful patterns to provide efficient extraction of hidden information. DM is the process of discovery huge amounts of data stored in the Data base. Large amount of data comes from different data sources and may be stored in various Data Warehouses; DM Techniques such as ML, AI and analytical modeling can be involved.

In Data Mining Process (Fig 1.1), it mainly focuses on:
1. Data Construction
2. Data Provision
3. Modeling

Shaik Johny Basha, Assistant Professor, Lakireddy Bali Reddy College of Engineering (A), Mylavaram, Andhra Pradesh, India.
(Email: shaikhjanibasha@gmail.com)

M.V.V.S. Subrahmanyam, Assistant Professor, Sasi Institute of Technology and Engineering (A), Tadepalligudem, Andhra Pradesh, India.
(Email: subrahmanyam.mavuri@gmail.com)

Data Construction:

It is the huge data obtained for mining. In this stage, subsets have been involved for:
i. Combining the information from various sources
ii. Data mining organization must be given input of some parts of data
iii. Finally process the data to the data mining stage.

Data Provision:

The Data which was mined and visualizing the information to the end user.

Modeling:

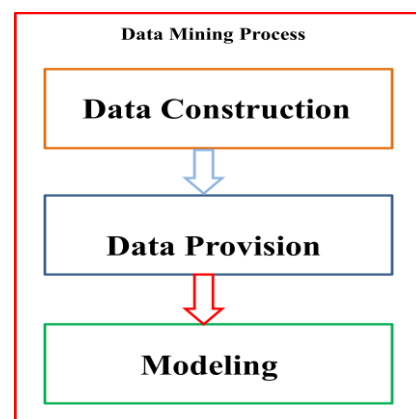It is a technique has to be selected to be used for the prepared data set.

Fig 1.1: Data Mining Process steps

II. LITERATURE REVIEW

Nowadays IoT & Data Mining applications are very demanding, exceptional to a distinctive enhance in the amount and density of data [6]. The IoT is an entirely a new set of challenges contain to the Data Mining domain [1] [2] [3]. The IoT can also be considering as a Universal System, which allows the communication between the Person to Person, Person to Things, which anything in the world by providing distinctive characteristics to each and every object [2]. Internet of Things is growing and continues to be the latest, most overvalued concept in the IT world.

S. No.	Author	Title	Year	Challenges
1	D'Angello, G. Ferretti	Multi-Level Simulation of IoT on Smart Territories	2016	Huge number of involved sensors and devices and heterogeneous scenarios
2	Hefnawy. A & Cherifi. C	Life Cycle Management in the Smart City Context	2016	High Complexity of Modern City Operations
3	Hefnawy. A & Cherifi. C, Bouras	IoT for Smart City Services : Life Cycle Approach	2016	Heterogeneous Sensors and Devices
4	Guar Scotney. B & G. Parr	Smart City Architecture and its applications	2015	Increased amount of Data
5	Mitton. N & Trivedi. K.S	Combining Cloud and Sensors in a Smart City Environment	2015	Differences in the numerous applications

III. IOT APPLICATIONS

1. Smart Home:

In recent days clearly smart home are very popular IoT application on all considered channels. In software sector every month nearly 70,000 peoples searching "Smart Home" in the world.

2. Wearable:

It is very popular topic these days. Wearable's application is to construct personal computing into new innovative directions it is a very hot topic full range of new capabilities enveloping connectivity can bring.


3. Smart city:

the main aim of smart city to overcome traffic, waste management and city safety and ecological monitoring. IoT the mainly focus on part of smart city to solve traffic jamming problems, to decrease the noise and pollution.


4. Smart Grids:

It is very special one. This mainly focuses on Behavior of electricity suppliers and clients in a robotic style to improve the effectiveness, reliability, and financial side of electricity.
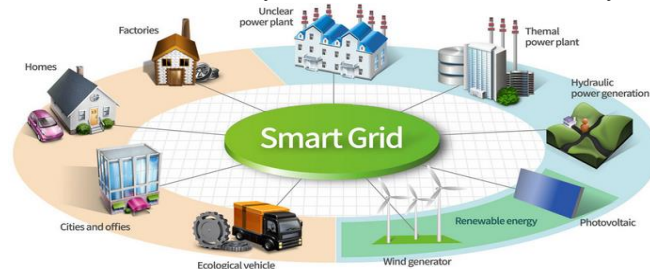
5. Smart Retail:

it is a foremost provider of customer insights and analytics that allow retailers to improve success. It allows the organization of products and services customized to person and various set of scientific solutions enabling us to change a predictable substantial store into an interactive point of sale.
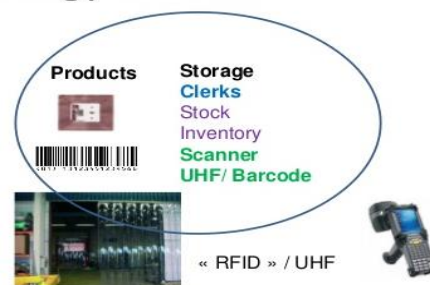
6. Smart Supply Chain:

It is organization disquiet the a different aspects of point to point logistics and provide sequence of management in the context of business, the IoT, AD Analytics and self-directed decisions enabled by AI.

IV. DIFFERENT CLUSTERING TECHNIQUES

In this section, we discuss the advantages and disadvantages of various data mining clustering techniques like K-means, Hierarchical, Parallel K-mean and DBSCAN.
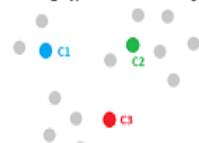
K-Means Algorithm:

It is one of the easiest and simplest techniques that solve the clustering problems in Data Analysis.

K-Means Algorithm
Step 1: Initialize Cluster Centers
Here We Select Cluster Points Randomly C1, C2, C3 & Represent with various colours such like Blue, Green, Red alone to signify the Cluster Midpoints.

Step2: Find the Distance among every various Data Points & Cluster Centers.
Step3: To allocate the Data Point to the Cluster Midpoints whose Distance from the Cluster Center is least of all the Cluster Midpoints
Step4: yet again estimate the new Cluster Center using

$$v_i = (1/c_i) \sum_{j=1}^{c_i} x_i$$

where, 'ci' represent the No. of Data Points in "ith" cluster,
xi set of data points,
vi set of centers.
Step5: Over again find the distance among every points and New Obtain Cluster Midpoints.
Step6: The condition no Data Point was reassigned then stop, Otherwise repeat from Step3.

DBSCAN Technique:

It is mostly used for Density Reachability & Density Connectivity. To do DBSCAN, we must need mainly two requirements:
1. eps
2. The minimum number of points required to form a cluster (minPts).

Analysis of IRIS Dataset:

We performed analysis on various iris data sets using DBSCAN Technique.
The main content of the IRIS Dataset consists of:
1. Sepal.Length
2. Sepal.Width
3. Petal.Length
4. Petal.Width
5. Species

Results:

We had shown the results of **K-Means Algorithm** & **DBSCAN Technique** applied on *iris* Dataset in Fig 1.2, 1.3, 1.4 and 1.5 respectively.
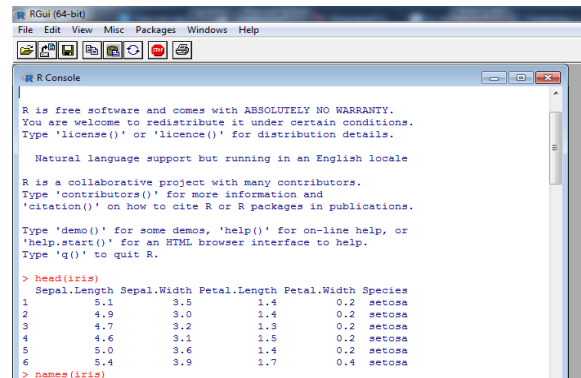

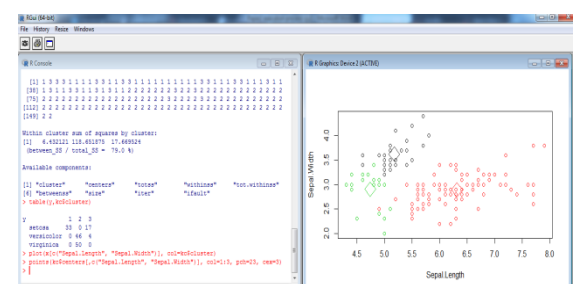
Fig 1.2: IRIS Dataset

Fig 1.3: k-means on IRIS Dataset
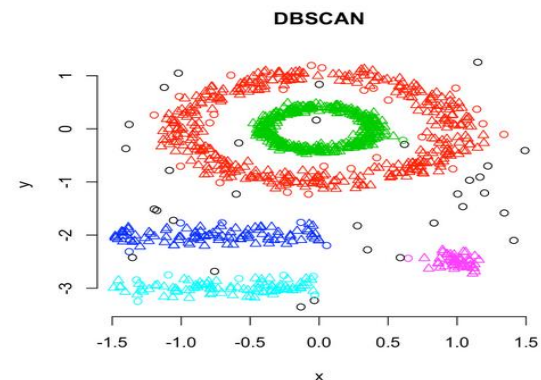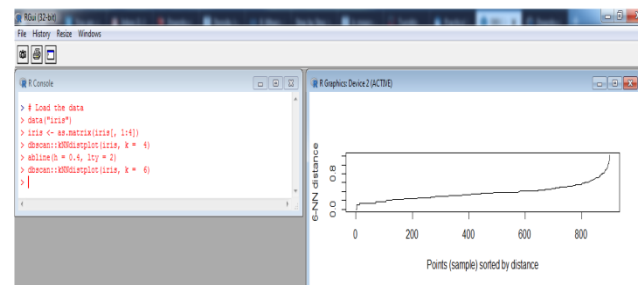

Fig 1.4: DBSCAN Technique on IRIS Dataset

Fig 1.5: DBSCAN Technique Analysis

Review on various techniques:

S. No.	DMT (Data Mining Technique)	Advantage	Disadvantage
1	k-means	Quite faster than other clustering techniques	Slower for Noise Datasets
2	Hierarchical	Gives a huge support for hierarchical nature between the different clusters	Cannot be undone the already done process
3	Parallel k-mean	Best for finding the independent distances	Gives Communication Overhead
4	DBSCAN	Much needed at Multidimensional Databases	Not suitable for sensitive and small databases

Fig 1.6: Review on Different Clustering Techniques

V. CONCLUSION

The Internet of Things shows novel group of facts mostly collected from various antenna strategies. In this article, we have given study of different data mining techniques and IoT applications. And we conclude that different clustering [6] and classification approaches. We also discussed about advantages & disadvantages few techniques as shown in the table format 1.6, We plan to perform a complete study on various IoT& DM Techniques apply various data sets in future work.

REFERENCES

1. Chen F, Deng P, Wan J, Zhang D, Vasilakos A, Rong X. Data Mining for the Internet of Things: Literature Review and Challenges.IJDSN. 2015;2015:1-14.
2. Ngai E, Xiu L, Chau D. Application of data mining techniques in customer relationship management: A literature review and classification.Expert Systems with Applications. 2009;36(2):2592-2602
3. S. Cuomo, P. D. Michele, A. Galletti and F. Piccialli, "A Cultural Heritage Case Study of Visitor Experiences Shared on a Social Network,"2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Krakow, 2015, pp. 539-544.
4. M&M Research Group, "Internet of Things (IoT) & M2M communicationmarket - advanced technologies, future cities & adoptiontrends, roadmaps & worldwide forecasts 2012-2017," Electronics.caPublications, Tech. Rep., 2012.
5. Angelo Chianese, Fiammetta Marulli, Francesco Piccialli, Paolo Benedusi, Jai E. Jung, An associative engines based approach supportingcollaborative analytics in the Internet of cultural things, Future Generation Computer Systems, 2016.
6. When Hu & Qing he Pan, "Data clustering and analyzingtechniques using hierarchical clustering method", Published in Springer Multimed Tools Appl (2015) 74: 8495. Doi: 10.1007/s11042-013-1611-9.