IJCSE International Journal of Computer Sciences and Engineering Open Access
Review Paper Vol.-7, Special Issue, 4, Feb 2019 E-ISSN: 2347-2693

A Review of Datamining Techniques in Internet of Things

G. Saranya[1*], M. Subashini [2]

[1,2]Dept. of Computer Science, R.A.College for women, Thiruvarur, Tamilnadu, India

Corresponding Author: saranya_tvr@rediffmail.com

Available online at: www.ijcseonline.org

***Abstract*—**We are entering in a new era of computing technology i.e Internet of Things(IoT). IoT is a network in which all physical objects are connected to the internet through network devices or routers and exchange data. IoT allows objects to be controlled remotely across existing network infrastructure. This technique also has autonomous control feature by which any device can control without any human interaction. The enormous amount of data produced by IoT devices can be converted into knowledge using data mining techniques. Data mining will play important role in constructing smart system that provides convenient services. It is required to extract data and knowledge from the connected things. For this purpose, various data mining techniques are used. Various algorithms such as classification, clustering, association rule etc. helps to mine data. This paper represents the different Data mining techniques with IOT.

***Keywords*—**Internet of Things, Data mining.

I. INTRODUCTION

The Main aim of Internet of Things (IoT) is to build an superior network technology which automatically catch the requirements of users and will operate in that point of view. At first in 1999 The Massachusetts Institute of Technology (MIT) put forwarded the concept of IoT. IoT is described as data and things round the clock connected through the Internet. The IoT also views everything as the same, as things. These things includes Smart phones, users, processing units tablets, Bluetooth, ZigBee, data centers (DCs), the Infrared Data Association (IrDA), cellular networks, ultra-wideband (UWB), near field communication (NFC) DCs, Wi-Fi networks, RFID, chips and sensors, vehicles, wristwatches, household equipments means IoT is mixture of factual things and virtual things which are connected anytime and anywhere. It is expected to touch 50 billion by the end of 2020.The data generated or collected by these devices is huge in volume. This data will be in vast amounts for a system, and immense for a larger systems. To conserve and produce significant business information out of this data, to provide different services to improvement of the business growth and system planning data mining is essential. A typical IoT system described in below in Figure 1

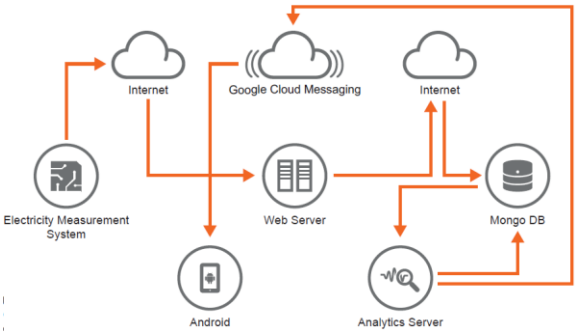

Fig. 1. Typical IoT System architecture

The main challenge is to select a suitable algorithm for a certain IoT system, as there several process and algorithms available in data mining.

The main purpose of any data mining technique is to set up a well-organized predictive or descriptive representation of vast data that not only best fits or describes it, even so able to generalize to newly generated data.

Data mining mainly categorized into two processes. One is Descriptive data mining and the other is Predictive data mining. In descriptive data mining data is described in a brief and aggregated way and gives significant general properties of the data. In predictive data mining data is analyzed in a sequence to build a single or a set of data models and endeavors to predict the behavior of newly generated data sets

by using techniques like regression, classification and trend analysis. Data mining can be viewed as an essential method in the course of knowledge discovery. This process is an iterative sequence of the following steps:

- *Data cleaning* - In this sequence noise and inconsistent data is removed.
- *Data integration* - This step combines multiple data sources.
- *Data selection* – In this data retrieved which is relevant to the analysis process.
- *Data transformation* – In this step summary or aggregation operations are performed so that data is transformed or consolidated into appropriate for data extracting.
- *Data mining* – In this process desired data patterns are extracted by using intelligent methods.
- *Pattern evaluation* - In this process discovery of interesting patterns that represents the knowledge depending on some attractive measures.
- *Knowledge presentation* - In this process mined knowledge is presented to the user by using visualization and knowledge representation techniques.

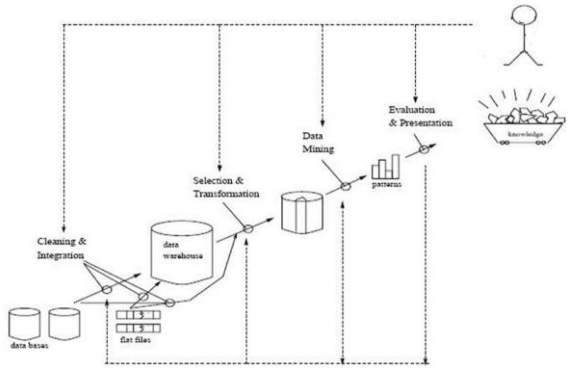

Figure 2: Data mining Architecture

IoT gathers data from different places, again that data may contain required data for IoT aside. Data collected by IoT is converted into serviceable information when we apply data mining to IoT, and then this information is converted into knowledge needed by user.

II. DATA MINING FRAMEWORK FOR IOT

Data mining is one of the most powerful and emerging technologies which is used to mine certain useful trends and patterns which are unknown, to enhance the performance of the organizations. Almost all the organizations are growing rapidly with the help of data mining functionalities Data mining helps in finding out something in the massive data which is unknown and most profitable to the organization. For example, by finding out the frequent buying patterns of the customers, a company can increase sales by placing those items together which are being purchased together, by applying discounts on those items or by reducing redundant items.

The primary aim of knowledge discovery in databases is to discover the novel patterns in the large data sets. It is the blend of many different domains which includes artificial intelligence, machine learning and statics. Data mining transforms a data set into understandable structure and extract important information which helps in gaining an insight into the raw data collected from various IoT applications.

Therefore, IoT forms a network of physical objects or things which are embedded with electronics, sensor and network connectivity by which the devices can collect and exchange data. The perfect association between IoT and Data mining results into a new innovative technology which will benefit every section of the society. This association give birth to a number of different applications. These applications generates enormous amount of heterogeneous data. As the data in IoT application is generated continuously from different sources like wireless sensor networks, RFID etc.,

Based on the data mining and IOT overview, the data mining in IoT process is shown in the following figure 3.

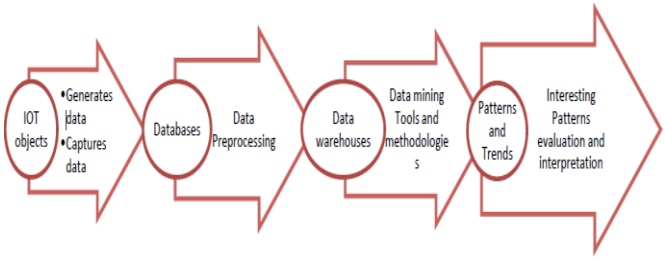

Fig. 3 Data mining framework for IoT

III. TECHNIQUES OF DATA MINING

To analyze large amount of data, data mining came into picture and is also called as KDD process. Data mining functionalities include classification, clustering, association analysis and prediction.

1. Classification

Classification is one of the data mining techniques which is useful for predicting group membership for data instances. Classification is data mining process that represents data items as labeled classes. With this process we can find the category of particular item in a dataset. For example if we consider the case of an automobile company we can analyze and predict the type of car customer wants to buy based on customer's profile ,family background and age by using classification model.

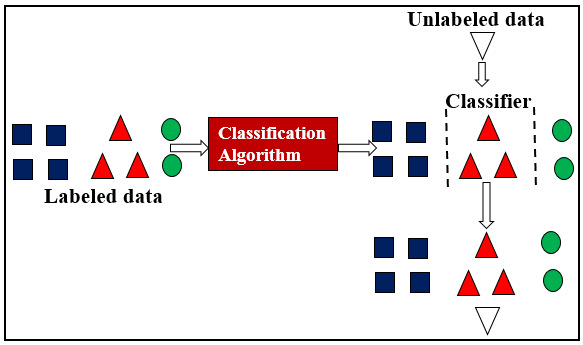

Fig 4: Classification process

There are several methods in Classification process in data mining are as follows:

- **Decision tree induction**: From the class labeled tuples the decision tree is build. Decision tree is tree like structure in which there are internal node, branch and leaf node. Internal node specifies the test on attribute, branch represents the outcome of the test and leaf node represents the class label. Two steps that are learning and testing are simple and fast. The main goal is to predict the output for continuous attribute but decision tree is less appropriate for estimating tasks. There may be errors in predicting the classes by using decision tree approach. Pruning algorithms are expensive and building decision tree is also an expensive task as at each level there is splitting of node.
- **Classification by back propagation:**Backpropagation is a neural network learning algorithm. Neural network learning is often called connectionist learning as it builds connections. It is feasible for that application where long times training is required. The most popular neural network algorithm is backpropagation. This algorithm proceeds in the way that it iteratively performs processing of data and it learns by comparing the results with the target value given earlier.
- **Lazy learners:** Eager learner is the form in which generalization model is being developed earlier before new tuple is being received for classifying. In lazy learner approach when given a training tuple it simply stores it and waits until a test tuple is given.It supports incremental learning. Some of the examples of lazy learner are K-nearest neighbor classifier and case- based reasoning classifiers.

2. Clustering

Clustering algorithms divide data into meaningful groups so that patterns in the same group are similar in some sense and patterns in different group are dissimilar in the same sense. Here, we do not have a training data. This is unsupervised learning technique. E.g., search engine uses clustering method to group several web pages into different groups like news, videos, images, blogs etc.

Clustering is an efficient way to enhance the performance of IoT on the integration of identification, sensing, and actuation. That is why many new clustering algorithms are developed for the WSNs that are probably the most common devices to be found on the IoT.

The clustering process includes various steps and it is a step by step process in which the results can be verified.

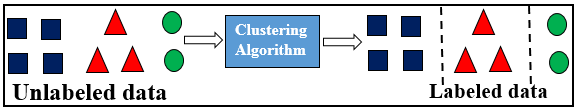

Fig 5: Clustering process

The main four steps followed are as below:

a) Feature selection or extraction: feature selection is selecting distinguishing feature form set of candidates and extracting means which it utilizes in the transformation to generate the useful and novel features from original ones .

b) Clustering algorithm design:

Every clustering algorithm is affected by measures. Next is to optimize the clustering solutions. As said by J. Klienberg that " It has been very difficult to develop a unified framework for reasoning about it (clustering) at a technical level, and profoundly diverse approaches to clustering".

c) Validation:

Validations of clusters are in the sense whether the groups formed are valid or not, the data is correctly identified according to groups. These all can be checked by main three indices which are known as testing criteria and these are as follows:

i. External indices
ii. Internal indices
iii. Relative indices

These indices are defined on different clustering structures that are known as partitioning clustering, hierarchal clustering and individual clusters.

d) Result interpretation: Next step is to provide accuracy to user and provide a meaningful insight form original data so that efficient results can be provided.

Methods of clustering

There are various methods for clustering which act as a general strategy to solve the problem and to complete this, an instance of method is used called as algorithm. Broadly clustering methods can be divided into two main categories which have number of instances. On the basis of that we have hierarchical and partitioning based methods. In hierarchical based clustering, the data sets of n elements are divided into hierarchy of groups which has tree like structure.

In partitioning based methods the output is like k partitions of N dataset elements

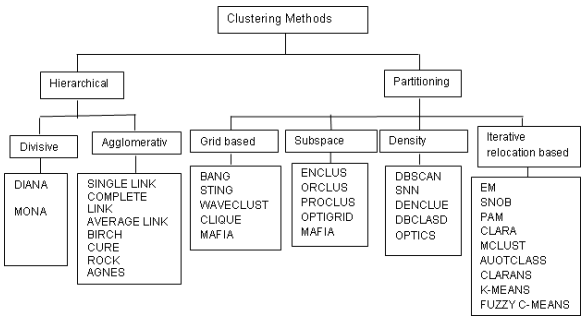

Fig.6 Categorization of clustering methods and algorithms

3. Association Analysis.

Association rule mining focus on the market basket analysis or transaction data analysis, and it targets to discover of rules showing attribute–value associations that occur frequently and also help in the generation of more general and qualitative knowledge which in turn helps in decision making . The research structure of association analysis is shown in Figure 6.

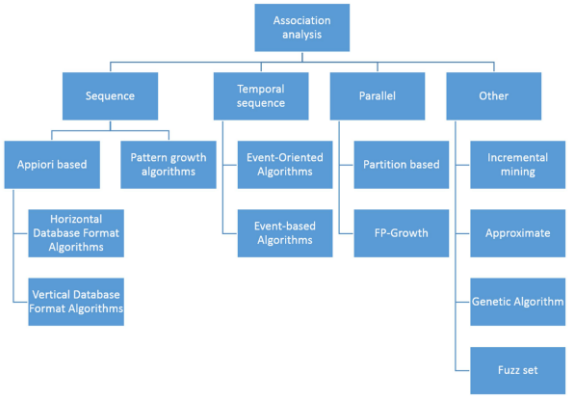

Fig 7: The research structure of association analysis

4. Prediction:

Prediction is relates with time series but not time bound. It is used to predict value based on past data and current data. (e.g) Water flow of a river will be calculated by various monitors at different levels and different time intervals. It then using those information to predict the water flow of future. Prediction is a wide topic and runs from predicting the failure of components or machinery, to identifying fraud and even the prediction of company profits. Used in combination with the other data mining techniques, prediction involves analyzing trends, classification, pattern matching, and relation. By analyzing past events or instances, you can make a prediction about an event.

IV. CONCLUSION

IoT generate enormous amount of heterogenous valuable data.To deal simultaneously with people and IoT devices, there arises the need of data mining techniques. 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.Adding more to the introduction of IoT and data mining, this paper focuses on techniques of Data mining.

REFERENCES

[1] H. Jiawei, M. Kamber, Data mining: concepts and techniques, Morgan Kaufmann Publishers, 2011.

[2] 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.

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

[4] 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, 2014.

[5] AsmitaGorave, VrushaliKulkarni, "Discrimination Aware Data Mining In Internet Of Things", Ijca, 0975 – 8887, Volume 159, February 2017.

[6] KrushikaTapedia, AnuragManoharWagh, "*Data Mining For Various Internets Of Things Applications*", Ijrat, E-Issn: 2321-9637, Ncpci-2016.

[7] 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.

