IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Survey on Data Mining for Internet of Things

Priyanka T. Avhad

Department of Computer Science and Engineering, MGM's JNEC Aurangabad, India

Abstract: Internet of Things (IoT) has been growing rapidly due to recent advancements in communications and sensor technologies. IoT has been playing an important role from traditional equipment to general household objects and has been attracting the eye of researchers from academia, industry, and government in recent years. Interfacing an every object together through internet looks very difficult, but within a frame of your time Internet of Things will drastically change our life. The big data captured by IoT are considered of high business also as social values and extracting hidden information from information, various data processing algorithms are often applied in Internet of Things (IOT) data. This paper gives idea and systematic review of various data mining processes and Algorithms also applications of data mining in Internet of Thing (IoT).

Keywords—Internet of Things,Data Mining,Application of Data Mining

I. INTRODUCTION

The Internet of Things (IoT) refers to the sort of the network which connect anything i.e. physical objects-devices, buildings, vehicles and other items embedded with software, sensors and network connectivity supported stipulated protocols that allows these objects to gather and exchange data. the Internet of Things(IoT) and its relevant technologies can seamlessly integrate classical networks with networked instruments and devices. IoT has been playing a crucial role ever since it appeared, which covers from traditional equipment to general household objects [1] and has been attracting the attention of researchers from academia, industry, and government in recent years.

Data mining is the process of extracting useful information or patterns from the raw data. Suppose we have a certain amount of data, and we look for events of a certain type within that data. Data mining involves discovering novel, interesting, and potentially useful patterns from large data sets and applying algorithms to the extraction of hidden information. Many other terms are used for data mining, for example, knowledge. discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging,and information harvesting [2]. The objective of any data mining process is to build an efficient predictive or descriptive model of a large amount of data that not only best fits or explains it, but is also able to generalize to new data [3]. On the basis of the definition of data mining and the Definition of data mining functions, a typical data mining process includes the following steps

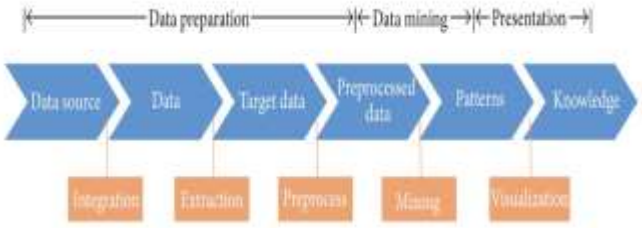

Figure1. The data mining overview

Data preparation: This step prepares the data for mining. It includes three different sub steps:

a) Integrate data in various data sources and clean the noise from data

b) Extract some parts of data into data mining system;

c) Preprocess the data to facilitate the data mining.

(ii) Data mining: This step involves to apply algorithms to the data to find the patterns and evaluate patterns of discovered knowledge.

(iii) Data presentation: It consist of visualization of the data and represent mined knowledge to the user.

II. RELATED WORK

Since Internet of Things is a completely new and one of the important concept, many researches are still at the initial stage. Right now, there are few works regarding data mining in the IoT.

Rob Petersen [4] come up with a new system CALLED ROAM, which identify inconsistency in moving objects. SpatioTemporal Sensor Graphs (STSG) proposed by Betsy

Aashi singh and Shilpi sharma [5] introduces the establishment of the relationship between Data Mining and the Internet of Things.There are Four models for data mining have been discussed indicating some challenges of the data for Internet of Things which are multi-layer model, distributed model, grid based model and big data mode

Vangipuram Radhakrishna [6] essentially focuses on design of novel similarity measure to find similarity between temporal patterns subjected to subset specifications which include specifying reference, a distance value permissible This research makes attempt to come up with an efficient similarity measure for revealing similar temporal patterns in time series data generated by IoT.

In [7], author proposed a design for high-performance data processing module of KDD for IoT with the three key considerations i.e. choosing objective, characteristics of data, and mining algorithm.

Objective: The relevant mining techniques has to be decided for the problem to be settled by the KDD. The suppositions, restrictions, and estimations of the problem should be find out first in order to accurately characterize the problem to be comprehended. With this data, the goal of the problem will be influenced precious stone to clear.
Data: Another imperative worry of data mining is the characteristics of data, for example, size, distribution, and representation. Distinctive data typically should be processed in a different way. In spite of the fact that data originating from various issues might be alike, they may must be investigated distinctively if the implications of them are unique.
Mining algorithm: Having above two parameters decided accurately, determining and selecting a data mining algorithm that suits to accomplish users task is incredibly much easier task.

In [7], author discussed three parameters which are vital to decide whether to develop new data mining algorithm or to use already designed algorithms. For an example, considering a scenario if we go to a derivation, that size and complexity of data that is needed to be processed is quite high that are beyond available system capabilities to process and no other options or techniques are available to minimize size and complexity of information then it is speculated to be solved using novel mining algorithm.

III.DATA MINING FOR IOT

A.Suitable Data Mining Processes for IoT

We are in a world where the speed with which the business must move is way faster than the time it takes to conceive and launch new solutions within the areas of huge data, data mining, cloud, and IoT [8]. To seek out relatively small part of data in peta byte sized databases generated from an IoT system is like searching for a black cat in a coal cellar. To urge in the game, different form of data mining algorithms should be built with various capabilities to urge insights and minimize the risk of project failures. Till today there are many studies which are trying to solve the matter of acquiring of huge data on IoT systems. Most of the mining techniques are developed to execute on a single system, so these KDD systems cannot be applied on to process big data of the IoT system, whereas for a small system undoubtedly these KDD processes will be applied directly. To develop a high geared data mining processing structure of KDD for an IoT system the subsequent three points [7] are to be considered to elect the suitable mining technology, and that they are

First and the foremost it is important to clear the definition of the given problem, also their limitations and all the required information.

Another part is to understand which kind of data will be required like the representation, size of data, and the processing of different data etc.,

• And at the end, on the basis of all the mentioned points, a suitable data mining algorithm is to be select to bring out sensible and required information from the raw data.

B. Data Mining Algorithms

Classification
It is defined a function of data mining that delegates items into categorical labels. It helps us to predict the category of selected item during dataset. Let's consider a scenario where a marketing manager of an automobile company wants to analyze the probability of a customer buying a kind of car by the idea of person's profile. A classification model is utilized to predict the type of car; family, sports, truck or van, that a customer is probably going to buy on the basis of person's age and family background [9]

There are different kind of classification models such as decision tree, neural networks, IF - THEN rules and it totally
Depends upon their use.

Clustering
Clustering is defined as categorizing the data into some sensible, meaningful groups or classes. It helps to obtain an easy method for the all users by grouping naturally. The example is a search engine which is based on clustering, that can be categorize from endless web pages to news, images, videos, reviews etc., There are different types of clustering models such as, Density based clustering, k Means clustering, Hierarchical clustering that can be used based upon their use

Association Analysis
Market basket is the best example for association. Market basket analysis is observed continuously in supermarket chains where the items or products which are likely to be bought together with another set of items are always placed together such as toothbrush and toothpaste are always in the same section. This helps in decision making. At first the given data is processed incessantly, for first catalog of association analysis. Apriori Algorithm is used to discover inter transactional association has been used followed up with association discovery. Other algorithms used are pattern growth, event oriented, event-based, partition based, FP Growth, Fuzzy set and incremental mining.

Time Series Analysis

When data points are present in consecutive time interval, time series analysis is applied to extract meaningful related to specific patterns or statistics. Stock market index value is analyzed in a time series manner. Time series analysis is also used in forecasting, to analyze dependent events; that is to predict future values based on past events. • Outlier Detection: Intermittently there exists a data which is not complaisant with general behavior or model of the data. This kind of data is different from remaining set of data which is called as outlier. This type of data contains useful information regarding aberrant behaviour of the system comprised of outliers. Outlier analysis can be used to extrapolate outliers, to calculate distance among objects, distribution of input space. The above mentioned data mining functionalities with the listed algorithms are the most commonly used algorithms in any field to mine the data and extract the required information.

IV. Application Of Iot With Data Mining

Since filleting advancement in IoT devices and sensors connected over internet, we've got a valuable set of applications with adaptation data processing during this technology. These applications include Automation of day to day smart home activities, smart city applications like traffic management, disaster management information, health care and vehicular applications like smart navigation and driverless driving etc.
Few of these applications are mentioned below

A. Smart City Applications

There are several IoT systems in smart city that relating to the appropriate data mining functionalities used to make existing system smarter and finer [10].

Traffic Control - IoT devices or things such as smart phones, vehicle sensors, GPS are employed all over the city can served as data points such as time of traveling, incidence of heavy vehicles, accident prone zones and construction areas.
Awareness of the causes behind traffic obstruction in the selected area is known with the help of these data points. To solve the traffic blockage problem we can use classification algorithm. The selected areas can be classified based upon the higher, lower or medium chances of traffic jam incidents in a specified location. For the prediction of the time in a day where the obstruction will happen at the highest rate and possible route that has no traffic blockage used to arrive at the destination is achieved by Classification method. To avoid congestion, classification will disperse the traffic

Suburban Electronic Meters - Traditional meters are replacing with smart electronic meters with a rapid pace, since smart meters can offer detailed description about real time energy consumption information in a digital way via e-mail or smart phones. Nonetheless, Time Series analysis is applied on time series data that is automatically collected at different time intervals all over the day and can be used to predict energy consumption, sends messages immediately if any anomaly is detected in energy consumption. Synthetic data which is used for forecasting may be generated available real data.

Pipeline Leak Detection – For municipal corporations, maintenance of water pipe leakages is burdensome more precisely with old pipes. Using outlier detection algorithm along with use of sensors, sound of water movement can be analyzed to spot leaks.

B. Smart Home Applications

For Governments IoT offers

Smart city applications: like power and lighting, surveillance, parking meter, adaptive traffic management, disaster management, events control and resource management ,
Smart transportation applications: like connected cars, fleet management, roadways, rail transport.

Smart grid applications: like power line efficiency, demand response
Smart water applications: like waste water management and domestic waterworks
Smart infrastructure and Environment based applications : such as air quality, environmental monitoring, landfill and waste management.

IoT devices generate data, by mining this data we can generate meaningful patterns which can further be used to predict future incidents for automated user interaction. This is obtain with classification and time series analysis models to classify the interactive devices that are connected together based on their utilization. Data produced by these devices can be stored with their relative time periods, by applying linear regression on this data can forecast future events

C. Enterprise Applications

On the side of enterprise IoT offers services like Energy based applications, Smart healthcare applications , Smart retails ,Smart agriculture applications , Smart banking applications, Smart building and Smart construction, Smart education , Smart insurance, Smart logistics and Smart manufacturing applications.
Energy based applications: such as operating management, as rigs and wells predictive maintenance, spill accident management
Smart healthcare applications: emergency ambulance service hospital management, emergency room, clinic based, surgery, lab diagnosis, home care, research, elder care, billing equipment efficiency
Smart banking application: ATM machine maintenance, online car, or home loans and e-statement

D. Health Care Applications

With the advancements of IoT systems services in health care industry are apparently enhanced. Health Care system in collaboration with IoT systems provides a numerous services for patients like frequent monitoring the degree pressure, diabetes, heart rate, weight details and pulse related information. of these data are stored the on cloud maintained by concerned hospital. An intelligent system should be developed to integrate these divergent data and supply accurately precise details about patient. With the assistance of text mining we will examine patient's medical record based ob doctor's prescription and conclude about the condition of patient. Clustering may be used for treatment of the patient

V. Conclusion and Future Scope

The Internet of Things concept rises with need to manage, automate, and explore all devices, instruments, and sensors in the world. To make wise decisions for people as well as for the things in IoT, data mining technologies are integrated with IoT technologies for decision making support and system optimization. However we must pay attention to think and efforts needed to IoT data mining systems.
\

REFERENCES

[1] Q. Jing, A. V. Vasilakos, J. Wan, J. Lu, and D. Qiu, "Security of the internet of things: perspectives and challenges," Wireless Networks, vol. 20, no. 8, pp. 2481–2501, 2014.

[2] H. Jiawei and M. Kamber, Data Mining: Concepts and Techniques, Morgan Kaufmann, 2011.

[3] A. Mukhopadhyay, U. Maulik, S. Bandyopadhyay, and C. A.C. Coello, "A survey of multiobjective evolutionary algorithms for data mining: part I," IEEE Transactions on Evolutionary Computation, vol. 18, no. 1, pp. 4–19, 2014

[4] Rob Petersen. "20 companies do data mining and make their business better". 2016

[5] Aashi Singh and Shilpi Sharma. Analysis on data mining models for internet of things. In I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC), 2017 International Conference on, pages 94–100. IEEE, 2017.

[6] Vangipuram Radhakrishna, Puligadda Veereswara Kumar, Vinjamuri Janaki, and Shadi Aljawarneh. A computationally efficient approach for temporal pattern mining in iot. In Engineering & MIS (ICEMIS), International Conference on, pages 1–4. IEEE, 2016.

[7] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, Laurence T Yang, et al. Data mining for internet of things: A survey. IEEE Communications Surveys and Tutorials, 16(1):77–97, 2014.

[8] Krushika Tapedia, Anurag Manohar Wagh," Data Mining for Various Internets of Things Applications", International Journal of Trend in Research and Development, Volume 4(3), ISSN: 2394-933

[9]] S. Song, Analysis and acceleration of data mining algorithms on high performance reconfigurable computing platforms [Ph.D. thesis], Iowa State University, 2011

[10] Krushika Tapedia, Anurag Manohar Wagh, "Data Mining For Various Internets Of Things Applications", Ijrat, E-Issn: 2321-9637, Ncpci-2016.