IFERP

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 5, Issue 3, March 2018

# IoT Services for Data Mining

[1] Anuradha D.Biradar [2] Sunita S.Gawali, [3] Dr.K.Vikram

[1][2] Student of M.Tech, [3] Professor

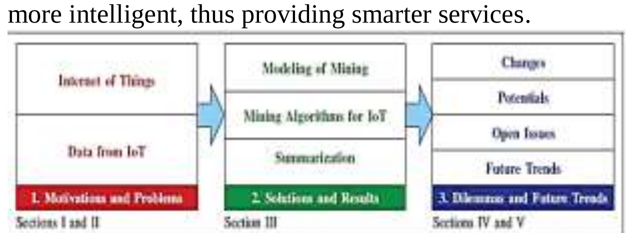
[1][2][3] Computer Engineering Department, G.H.Raisoni College of Engineering & Management, Wagholi, Pune

***Abstract:*** **- It sounds like mission impossible to connect everything on the earth together via internet, but Internet of Things (IoT) will dramatically change our life in the foreseeable future, by making many "impossibles" possible. To many, the massive data generated or captured by IoT are considered having highly useful and valuable information. Data mining will no doubt play a critical role in making this kind of system smart enough to provide more convenient services and environments. This paper begins with a discussion of the IoT. Then, a brief review of the features of "data from IoT" and "data mining for IoT' is given. Finally, changes, potentials, open issues, and future trends of this field are addressed.**

***Keywords:*** **Internet of Things, data mining.**

## I. INTRODUCTION

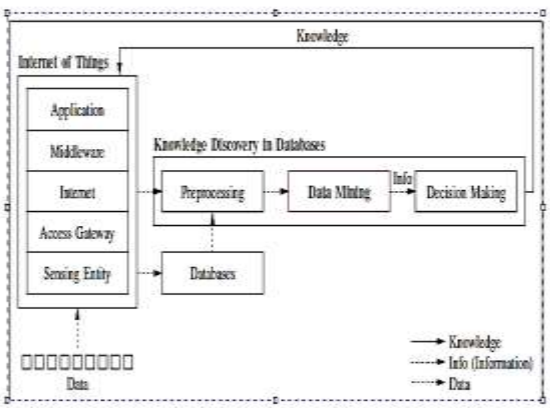
Since the first day it was conceived, the Internet of Things (IoT) [1], [2], [3] has been considered the technology for seamlessly integrating classical networks and networked objects [4]. The basic idea of IoT is to connect all things1 in the world to the internet. It is expected that things can be identified automatically, can communicate with each other, and can even make decisions by themselves. In fact, it is even anticipated that the emergence of IoT as a service provider will be a trend although not all the technologies used are brand-new. In addition to the great progress on computer communication and relevant technologies that make many applications possible, a good news [5] is that IoT and machineto- machine were worth $44.0 billion in 2011 and are expected to grow up to $290.0 billion by 2017. Thus, researchers from different industries, research groups, academics, and government departments have paid attention to revolutionizing the internet, by constructing a more convenient environment that is composed of various intelligent systems, such as smart home, intelligent transportation, global supply chain logistics, and health care [3], [4], [6], [7]. To clarify what the IoT refers to, several good surveys were presented recently each of which view the IoT from a different perspective: challenges [4], applications [7], standards [8], and smartness [9], [10]. Among these surveys, a comprehensive and semantics, was presented by Atzori and his colleagues [3]. Another recent study [6], inspired by [3], presented a generic five-layer architecture to describe the overall design of IoT. From the bottom up, the five layers are: edge technology, access gateway, internet, middleware, and application. Besides describing the infrastructures and things of IoT, a number of more recent studies [11], [12], [13], [10], [4] emphasized that most things on the IoT are supposed to have intelligence, thus are called "smart objects" (SO) and are assumed capable of being identified, sensing events, interacting with others, and making decisions by themselves [13], [10]. One of the most important questions that arise now is, how do we convert the data generated or captured by IoT into knowledge to provide a more convenient environment to people? This is where knowledge discovery in databases (KDD) and data mining technologies come into play, for these technologies provide possible solutions to find out the information hidden in the data of IoT, which can be used to enhance the performance of the system or to improve the quality of services this new environment can provide. A numerous researches are therefore focusing on using or developing effective data mining technologies for the IoT. The results described in [14], [15], [16], [17] show that data mining algorithms can be used to make IoT more intelligent, thus providing smarter services.




Technically, all things on the IoT may create a data deluge that contains different kinds of valuable information. However, technical issues and challenges on how to handle these data and how to dig out the useful information have emerged in recent years. A simple taxonomy [18] to differentiate the types of data from IoT is to use "data about things" to refer to data that describe things themselves (e.g., state, location, identity, and so on) and "data generated by things" to refer to data generated or captured by things. Normally, the former

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

contains data that can be used to optimize the performance of the systems, infrastructures, and things of IoT whereas the latter contains data that are the results of interaction between humans, between human and systems, and between systems that can be used to enhance the services provided by IoT. divided into two steps: the data processing step (consisting of the selection, preprocessing, and transformation steps), which is to be taken before the data mining step, and the decision making step (consisting of the interpretation/evaluation step), which is to be taken after the data mining step.




## III. DATA MINING FOR IOT

The relationships between big data, KDD, and data mining for IoT will be discussed in this section. A simple model for determining the applicable mining technologies and a brief introduction to the well-known data mining technologies for IoT will also be given in this section, by using a unified data mining framework and a few simple examples. After that, a detailed analysis and summarization of mining technologies for the IoT will be given.

A. Basic Idea of Using Data Mining for IoT It is much easier to create data than to analyze data. The explosion of data will certainly become a serious problem of IoT. Until now, a numerous studies [14], [15], [16], [17] have attempted to solve the problem of inquiring big data on IoT. Without effective and efficient analysis tools, we, and all the systems, will definitely be submerged by this unprecedented amount of data. When KDD is applied to IoT, from the perspective of hardware, cloud computing and relevant distributed technologies are the possible solutions for big data; nevertheless, from the perspective of software, most mining technologies are designed and developed to run on a single system. In the circumstances of big data, it is almost certain that most KDD systems available today and most traditional mining algorithms cannot be applied directly to process the large amount of data of IoT. Generally speaking, either the preprocessing operator of KDD or the data mining technologies need to be redesigned for IoT that can produce a large amount of data. Otherwise, the data mining technologies today can only be applied to small scale IoT system that can produce nothing but a small amount of data. To develop a high-performance data mining module of KDD for IoT, the three key considerations in choosing the applicable mining technologies for the problem to be solved by the KDD technology—the objective, characteristics of data, and mining algorithm—are as given below. • Objective (O): The assumptions, limitations, and measurements of the problem need to be specified first so as to precisely define the problem to be solved. With this information, the objective of the problem can be made crystal clear. • Data (D): Another important concern of data mining is the characteristics of data, such as size, distribution, and representation. Different data usually need to be processed differently. Although data coming from different problems, say, Di and Dj , may be similar to each other, they may have to be analyzed differently if the meanings of the data are different. • Mining algorithm (A): With needs (objective) and data clearly specified above, data mining algorithm can be easily determined, as to be discussed in Section III-E2. Whether or not to develop a new mining algorithm can be easily justified by using these factors. For instance, from the characteristics of data, if the amount of data exceeds the capability of a system and if there is no feasible solution to reduce the complexity of the data, then a novel mining algorithm is definitely needed; otherwise, the current mining algorithm suffices. Another consideration is related to the property and objective of the problem itself. If a novel mining algorithm can enhance the performance of a system, then the new mining algorithm is also needed. An example is the clustering algorithm for a wireless sensor network, which needs to take into account the load of computation, but most traditional clustering algorithms simply ignore this issue. Now that the objective of the problem is decided, the characteristics of the input data are understood, and the particular goals of mining and the mining algorithms are chosen,

**Algorithm 1** Unified Data Mining Framework

```
Input data D
Initialize candidate solutions r
While the termination criterion is not met
    d = Scan(D)  [Optional]          S
    v = Construct(d, r, o)           C
    r = Update(v)                    U
End
Output rules r
```

**Algorithm 2** k-means algorithm

```
Input data D
Randomly create a set of centroids c
While the termination criterion is not met
    v = Assign(D, c)                 SC
    c = Update(v)                    U
End
Output centroids c
```

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 3, March 2018

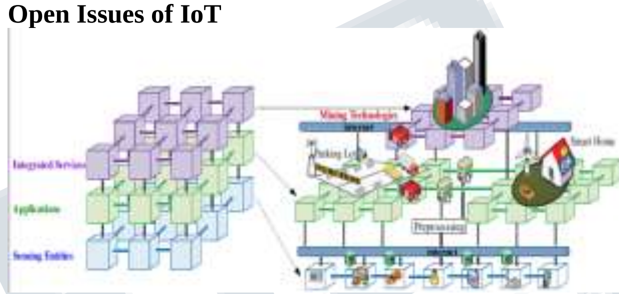
**Clustering for Services of IoT:**
Another research issue in using clustering technologies for the IoT is having services and applications of the IoT ake decisions by themselves or provide better services, such as detecting the falling event of older people by using a smart home system.

**Classification for Outdoor Services of IoT:**
Researches on using classification for the IoT can be divided into two categories: outdoor and indoor, depending on the region. One of the nightmares occurred frequently in our daily life is the traffic jam problem, especially when you are living in a big city. More and more researches therefore are focusing on the traffic jam problem, by using mobile devices or smart phones to interchange information to avoid getting stuck on road, such as traffic forecasting services [75] and accident detection [76]. Since getting information about the traffic situation via the internet and social network has become a trend, in [77], a driver guidance tool was developed by integrating the location information provided by the GPS, geographical information tracking from vehicles, and other information that can becollected from the internet to predict the future traffic situation. The decision tree classification algorithm is then used by this kind of system to predict and suggest the routing path for a driver, based on real-time information, historical data, and so on.

**Which Mining Technologies Are Applicable:**
Now that a unified framework has been presented to give a systematic description of data mining algorithms and how they are applied to IoT, the question that arises is, what mining algorithm is suitable for the development of a high-performance system or for the provisioning of a better service in different IoT environments that we may encounter. Fig. 6 gives a classification matrix to help us differentiate clustering, classification,

**Changes Caused by IoT**
In this section, changes for the IoT inspired by [3] are discussed from three different perspectives—thing-oriented, internet-oriented, and semantics-oriented The very first kind of change is on the "thing." A very simple way to differentiate a device is if it can automatically become part of the IoT; that is, if it can be automatically connected to the internet, either directly or indirectly through other devices disregarding its capability. Another change is that devices are made smaller and smaller, implying that the lifetime of battery can be prolonged and thus more functions can be integrated into each device, such as integrating RFID and wireless sensor into a single device. These changes make it possible for every thing to upload the data it collected to the server or application. We can then imagine that all the things on the planet being connected together [135] will come true in the foreseeable future. Owing to these changes on things, data mining technologies are now able to enhance the performance of devices, to give them basic autonomous or cognitive ability to make decisions by themselves (e.g., to give air-conditioner the autonomous ability to automatically adjust the temperature of a room) and to interact with other things (i.e., machine to machine).

**Open Issues of IoT**




Although the focus of mining problems for the IoT differs from that of the traditional mining problems, they still inheritmany of the open issues of the traditional mining algorithms and problems in the development of the mining algorithms for IoT, such as scalability for large-scale data set. There is no need to say that many other open issues need to be addressed in the design of the mining algorithms for IoT. With the technologies available today, here are some of the open issues:

## V. CONCLUSIONS

In this paper, we review studies on applying data mining technologies to the IoT, which consist of clustering, classification, and frequent patterns mining technologies, from the perspective of infrastructures and from the perspective of services. The analysis and discussions on the scale of each mining technology and of the overall integrated system are also given. To make it easier for the audience of the paper to fully understand the changes brought about by the IoT, a discussion, which goes from the changes caused by the IoT in using mining technologies to the potentials to the open issues we are facing nowadays, is presented.

## REFERENCES

[1] K. Ashton, "That 'Internet of Things' Thing," 2009, RFID Journal, Available at http://www.rfidjournal.com/article/print/4986.

[2] Auto-ID Labs, Massachusetts Institute of Technology, 2012, available at http://www.autoidlabs.org/.

[3] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey,” Computer Networks, vol. 54, no. 15, pp. 2787–2805, 2010.

[4] D. Miorandi, S. Sicari, F. De Pellegrini, and I. Chlamtac, “Internet of things: Vision, applications and research challenges,” Ad Hoc Networks, vol. 10, no. 7, pp. 1497–1516, 2012.

[5] M&M Research Group, “Internet of Things (IoT) & M 2M communication market - advanced technologies, future cities & adoption trends, roadmaps & worldwide forecasts 2012-2017,” Electronics.ca
Publications, Tech. Rep., 2012.

[6] D. Bandyopadhyay and J. Sen, “Internet of things: Applications and challenges in technology and standardization,” Wireless Personal Communications, vol. 58, no. 1, pp. 49–69, 2011

Anuradha D.Biradar is pursuing M.Tech in Computer Engineering at G.H. Raisoni College of Engineering & Management, Wagholi, Pune. Her intrest areas are Cloud computing and IoT, Network Securtiy and Image Processing. As a Student Level she published two international journals.

Sunita S.Gawali is pursuing M.Tech in Computer Engineering at G.H. Raisoni College of Engineering & Maqnagement,Wagholi,Pune , Her interest areas are Cloud computing and Computer Organization , Java, ,Python, Network Security and Image Processing. As a Student Level she published two international journals

Dr.K.Vikram working is a professor in G.H. Raisoni College of Engineering & Management. He published many research papers and published two text books like Digital Image processing and Computer organization and he is editorial board member of many international journals