A Research Direction on Data Mining with IOT

Conference Paper · March 2017

DOI: 10.1007/978-3-319-63673-3_22

4 authors, including:

Purvi Prajapati
Charotar University of Science and Technology
8 PUBLICATIONS **18** CITATIONS
SEE PROFILE

Jalpesh Vasa
Charotar University of Science and Technology
6 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Jay Patel
Charotar University of Science and Technology
1 PUBLICATION **0** CITATIONS
SEE PROFILE

A Research Direction on Data Mining with IOT

Foram Chovatiya[1], Purvi Prajapati[1], Jalpesh Vasa[1], Jay Patel[1]
Charotar University of Science and Technology,Gujarat, India
foramchovatiya@gmail.com, purviprajapati.it@charusat.ac.in
jalpeshvasa.it@gmail.com, jaypatel.it@charusat.ac.in

Abstract. The mission of connecting everything on the earth together via internet seems to be impossible. . There will be the great effect on human life by Internet of Things (IOT), because with the help of IOT, many impossible things will become possible. IOT devices generates big data having useful, valuable and highly accurate data. It is difficult to extract the required information or data from the set of big data discovered by any device. For this purpose, data mining is used. Data mining will plays 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 mining etc helps to mine data. This paper represents the different Data mining techniques, challenges, and Data mining issues with IOT.

Keywords: Data mining, Internet of things, clustering, classification, frequent pattern.

1. INTRODUCTION

IOT deals with connecting each and every things of the world via Internet [1,2,3]. The great progress on information technology and computer communication has made many application possible [4]. IOT is the next advanced generation of internet. It is believed that IOT will help in connecting trillion of nodes of various objects with the large web servers and cluster of a supercomputer. IOT also helps in integrating new computing technologies and communication [5]. Since, a decade, mobile devices and ubiquitous services help people to connect with anyone anywhere in the world. Nowadays, with the help of these devices, there is no limitation that corrupts the connection between people[6].Many researchers working in different fields like academics, institutes and government departments have shown keen interest in modifying the internet by designing various systems like smart home, smart pen, intelligent transportation, global supply chain, healthcare[3,7,8,9] etc.

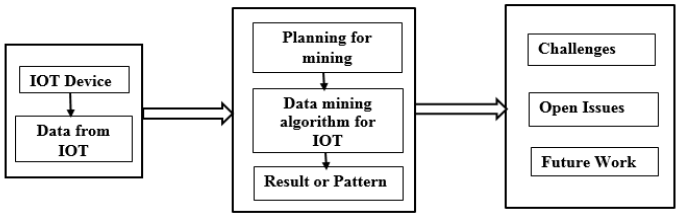

Fig 1. Roadmap of Data Mining with IOT [4]

Above Figure Shows the cycle of the process required to design the survey of data mining with IOT. During the initial phase of the work, we need to review all the information about IOT device. We will extract the data stored in that device. Now our next step is to design such a type of data mining algorithm which will help in connecting data mining with IOT. While connecting data mining with IOT we may face some challenges and open issues that are described in the later section of this paper [4].

2. DATA FROM IOT

It is obvious that IOT may create data consisting of much useful information. But, recently technical issues and challenges are major research topic along with different methods of handling these data. It is possible to solve the problem of[4] large IOT data is to design sensors capable of collecting only the interesting and useful data. To reduce the complexity of input data[4] is the recent research trend. Distributed computing, cloud computing and feature selection[4] are the popular data handling methods. It is assumed that big data will give born to more number of patterns from services and applications[4].

Moreover, how to handle large data obtained from IOT devices and finding hidden information from the data is important task. As described, data analysis for sensors and devices will be helpful in developing some useful system for the smart city or smart home. Many possible applications are possible to be developed from the large data analysis process. For fulfilling the task of finding hidden information from big data, Knowledge Discovery in Databases (KDD) successfully applied to different domains. Applied KDD has capability of finding "Something" or "interesting pattern" from IOT, with the help of following steps: data gathering, preprocessing, data mining and evaluation or decision making. Of these, data mining step play a main role in extracting interesting[4] pattern or knowledge from the big data. Data fusion, data transmission, big data and data decentralized issues has an impact on the system performance and IOT service quality[4].

3. DATA MINING FOR IOT

The next section describes relationship between big data and data mining for IOT and detailed analysis and summarization of different data mining techniques for the Internet of Things[4].

Basics of using data mining for IOT

It is easier [4] to use data mining for IOT for creating and analyzing data. Till the date, many studies have tried solving the problem of finding big data on IOT, without using effective and efficient analysis tools [4]. In recent times where big data is used widely, KDD systems and most traditional algorithms [4] are impossible to apply directly on the large amount of IOT data.

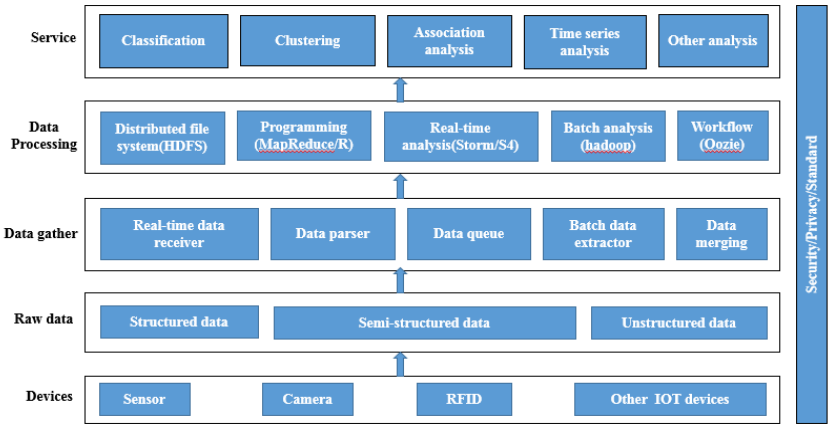

Fig 2. Big data mining system for IOT [14]

For developing a high-performance data module of KDD for IOT [4], the applicable mining technologies is to be solved by KDD process: - These are Objective, Characteristics of data, and mining algorithm [4].

Objective: It is important to specify the limitations, assumptions, measurement and issues of the problem to get it solved [4]. The objective of the problem can be made clear with the help of above described parameters [4].

Data: Characteristics of data such as distribution, representation, size [4] etc. plays a vital role in data mining. Various data normally need to be processed differently. For example, R_i and R_j may be similar or not, but they need to be analyze differently if the semantics of the data is various [4].

Mining Algorithm: Data mining algorithm [4] can be designed easily with the specified needs of objective and data application. Here the question arises that, which data mining algorithm is suitable for [4] increasing high performance of the system or for the finding of a better service in [4] various IOT environments [4].There are different techniques for extracting bigdata from IOT. These techniques are termed as Classification, Clustering and Association Rule Mining.

Classification: It is used to classify labeled as well as unlabeled patterns [4]. Labeled data basically considers the piece of unlabeled data with some information, class or tag related to it. For example, the picture of any animal is said to be the part of unlabeled data unless any information about it, its name, or tag like its sound is not clearly mentioned. Labeled data are achieved by making judgments about available piece of unlabeled data [4].

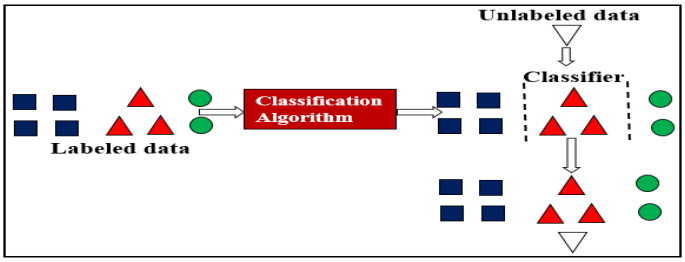

Fig 3.Classification process [4]

Clustering: It is used to classify unlabeled patterns [4].The example of unlabeled patterns are the samples of natural or human-created facts. In addition to this facts, the category of unlabeled patterns include photos, videos, audio recordings, x-rays, tweets, news articles, etc. There is no specific knowledge about the constituent of the unlabeled patterns, but contains only information of data.

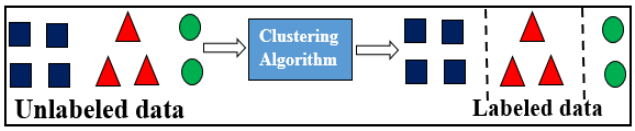

Fig 4. Clustering process[4]

Association: It is used to find an event from the input pattern that doesn't occurs in particular order [4]. Sequential pattern is a part of association rule mining and is used to find an event from the input pattern that occur in particular order [4].

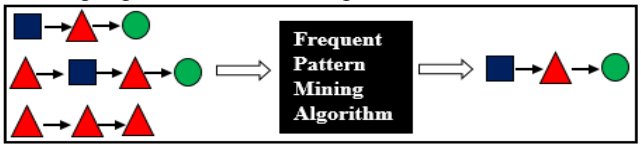

Fig 5.Frequent pattern mining process [4]

Here we can apply total 12 numbers of possible combinations to extract bigdata from IOT as described in the following table 1.

Table 1. Combination of Algorithm

Algorithm	Possible	Combination
Clustering	Clustering→Classification	Clustering→Classification→Frequent Pattern
	Clustering→Frequent Pattern	Clustering→Frequent Pattern→Classification
Classification	Classification→Clustering	Classification→Frequent Pattern→Clustering
	Classification→Frequent Pattern	Classification→Clustering→Frequent Pattern
Frequent Pattern	Frequent Pattern→Clustering	Frequent Pattern→Clustering→Classification
	Frequent Pattern→Classification	Frequent Pattern→Classification→Clustering

The objective of finding an interesting hidden pattern may be different depending upon the goal. Many researchers try to provide better services by combining different mining techniques. Overall study and system designing is needed because a single technique or algorithm will not work for extracting [4] useful information and making a decision. For this reason, we describe the possible combination of mining technologies.

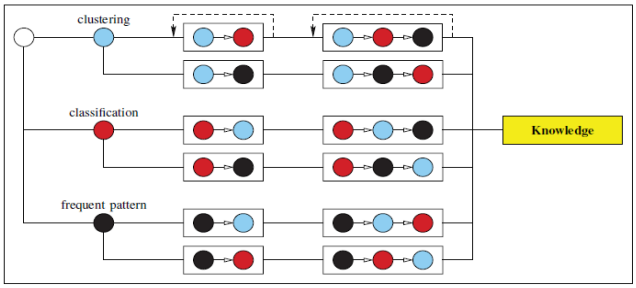

Fig 6.Different combination of mining technologies for the IOT [4]

Fig 6. Shows possible combination of mining technologies that could apply to extract hidden information. The first combination shows the use of clustering algorithm first and then classification algorithm. The second combination shows that classification algorithm to use first and after that clustering algorithm applied on the data. In the first combination, a set of classifiers is created without having any knowledge of input data or pattern. The second combination, classification is responsible for creating set and classifiers. On the set of the classifier, clustering algorithm used to add new classifier and create a new pattern. It is the combination of different mining technologies that enables the possibilities of handling patterns [4] or knowledge that enters the IOT incrementally.

For example, recognizing the human faces or behavior not previously in the knowledge database (the set of classifiers)[4]. The different combinations that can be used are, like

Clustering → classification → frequent pattern

The three techniques mentioned above consists of number of algorithms that are used to mine data from IOT devices. Some of them which are scalable and efficient for extracting data are decision tree algorithm (Classification technique), k mean algorithm (Clustering technique), and frequent pattern mining (Association rule mining).

It is mandatory to evaluate the quality of data that is obtained by applying one of the above mentioned algorithm. The best way to calculate and count the patterns in the right group is termed as "Accuracy Rate" (AR). It is defined by,

$$AR = \mathrm{Nc}/\mathrm{N_t} \qquad (1)$$

Where Nc is the total number of test patterns assigned correctly to the group they belong and N_t is the number of test patterns [10]. The other parameters which helps in measuring the quality of the accurate data are known as precision (P), and recall (R). The combination of P and R is termed as F-score which is defined as follows [10],

$$F = \frac{2\mathrm{PR}}{\mathrm{P+R}} \qquad (2)$$

4. CHALLENGES IN DATA MINING WITH IOT

1. A major challenge is to upgrade a crime detecting application which includes advanced features that prevent crime [11, 12].
2. One challenge is to extract large data available in large data storage and to detect any noise or unreliable data in that large dataset [11].
3. Mining uncertain and incomplete data is also a big data challenge. The algorithm modification is also very difficult, providing security solutions for sharing data is not so easy [11].
4. Conversion of IOT generate data into knowledge for providing a convenient environment to the user is a challenge [4,13].
5. Analysis and handling of large data is a difficult task for data mining purpose [4].
6. Selecting and implementing the better & efficient technology and mapping it with other technologies is difficult [4].
7. Building an intelligent system with the help of simple algorithm is not possible. For preparing an intelligent system, a number of algorithms must be fused to a single algorithm [4].
8. Internet connection may arise as a challenge when used in rapid speed device [4].
9. It is a challenge to build up an intelligent industrial IOT device, which includes smart city, green energy generation etc [14].
10. Major challenges in security are widespread data collection, unexpected use of consumer data [14].
11. Sharing of standards and infrastructure gives rise to a security issue [3].

5. MAJOR ISSUES IN DATA MINING WITH IOT

1. Parallel programming needs to be designed in such a way that every algorithm can be applied to it [11, 15].
2. The framework that considers security, privacy, data sharing, the growth of data size etc is to be designed [11].
3. Infrastructure Perspective: IOT gives low computation and high throughput, but mining algorithm is designed for small size & low power consumption device. This creates an infrastructure issue [4, 16].
4. Data Perspective: Gathering data from different sources creates redundant data. The user needs to filter the redundant data for better system performance. Moreover, data generate from different from sources may become an obstacle [4, 17].
5. Algorithm Perspective: Some example of IOT needs adding classifiers dynamically and some needs adding classifier statically. So, we need some mining technologies to be fused that can classify the classifiers in a common way [4].

6. Privacy and Security: Privacy and Security remains an issue because every algorithm and technologies are not able to outperform privacy and security issue [4, 18]. For example, it is easy for companies to collect different customer data from various devices or sources and use data mining techniques to find the information that[2] helps in increasing sales volume but the issue is that many customers wouldn't like to disclose their privacy and security, such as retail, shopping behavior.
7. Massive Scaling: How to name, identify, authenticate, maintain, protect and use a large number of data is an open issue [18, 19].
8. Architecture and dependency: It is difficult to construct an architecture that connects a large number of things to the internet. Many things are dependent on each other, so removing any of them may generate an error[18].
9. Robustness: In robustness, devices location have to be known if clock drifts, the location of the device may not be accurate [18].

6. CONCLUSION

The concept of IOT has taken born from the requirement of managing, automating and exploring all devices, instruments and sources all mover the world. To deal simultaneously with people and IOT devices, there arises the need of data mining techniques. Various techniques of IOT supports in decision making and optimization of any application. When fused with IOT, data mining basically deals with discovering useful and interesting patterns from large data. Then number of algorithms are applied to extract important hidden information. Data mining techniques includes clustering, classification, pattern mining. Adding more to the introduction of IOT and data mining, this paper focuses on challenges and open issues of data mining. Characteristics of big data, data mining algorithms and analysis on challenges is also described. At this stage when the development of IOT is at initial stage, focus is more on developing efficient preprocessing mechanism and consequent developing effective mining technologies for describing rules of IOT data. In the addition to research trends going on nowadays in the world are discussed in this paper. The research trends are: (i) Data will be uploaded the internet once connected, but IOT faces come problem such as, data abstraction, data summarization and data fusion[2]. (ii) To mine data from the application controlling multi-media devices and controlling energy are also a topic of research.

REFERENCES

1. Feng Chen, Pan Deng, Jiafu Wan, Daqiang Zhang, Athanasios V. Vasilakos, and Xiaohui Rong, "Data Mining for the Internet of Things: Literature Review and Challenges", Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, 2015.

2. Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang, "Data Mining for Internet of Things: A Survey", IEEE Communications Surveys and Tutorials, 2014.
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. John A. Stankovic, "Research Directions for the Internet of Things", IEEE, 2014.
5. Shen Bin, Liu Yuan, Wang Xiaoyi, "Research on Data Mining Models for the Internet of Things", IEEE, 2010.
6. Yan Chen, Ai-xia Han Cai-hua Zhang, "Research on Data Mining Model in the Internet of Things", International Conference on Automation, Mechanical Control and Computational Engineering, 2015.
7. H. Chen, W. Chung, J. J. Xu, G.Wang, Y. Qin, and M. Chau, "Crime data mining: a general framework and some examples," Computer, vol. 37, no. 4, pp. 50–56, 2004.8] X. Wu and S. Zhang, "Synthesizing high-frequency rules from different data sources," IEEE Transactions on Knowledge and Data Engineering, vol. 15, no. 2, pp. 353–367, 2003.
8. X. Wu and S. Zhang, "Synthesizing high-frequency rules from different data sources." IEEE Transactions on Knowledge and Data Engineering, vol. 15, no.2. pp. 353-367, 2003.
9. T. Keller, "Mining the internet of things: Detection of false-positive RFID tag reads using low-level reader data," Ph.D. dissertation, The University of St. Gallen, Germany, 2011.
10. A. de Saint-Exupery, "Internet of things strategic research roadmap," European Research Cluster on the Internet of Things, Tech. Rep., 2009.
11. R. A. Baeza-Yates and B. A. Ribeiro-Neto, Modern Information Retrieval. ACM Press Addison-Wesley, 1999.
12. T. He, J. Stankovic, C. Lu and T. Abdelzaher, "A Spatiotemporal Communication Protocol for Wireless Sensor Networks", IEEE Transactions on Parallel and Distributed Systems, Vol. 16, No. 10, Oct. 2005.
13. B. Liu,"Web Data Mining: Exploring Hyperlinks, contents, and Usage Data". Berlin, Heidelberg: Springer-Verlag, 2007.
14. K.Ashton, "That 'Internet of Things' Things,"2009, RFID Journal, Available at http://www.rfidjournal.com/article/print/4986.
15. Auto-ID Labs, Masschusetts Institute of Technology, 2012, Available at http://www.autoidlabs.org/.
16. L.Atzori, A. Iera, and G. Morabito,"The internet of things: A survey," Computer Networks, vol.54, no.15, pp. 2787-2805, 2010.
17. D. Miorandi, S.Sicari, F. DePellegrini, and I.Chlamtac,"Internet of things: Vision, applications and research challenges", Ad Hoc Networks, vol.10,no.7, pp.1497-1516,2012
18. 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.
19. M.C. Domingo, "An overview of the internet for things for people with disabilities," Journal of Network and Computer Applications, vol.35, no,2, pp.584-596, 2012.