A Review on Data Mining Techniques for Internet of Things

Tharushika Samarasingha[1], Manage Prabhashrini Dhanushika[2]

[1,2]Department of Computational Mathematics, University of Moratuwa, Sri Lanka
[1]tharushikasupipi@gmail.com, [2]dhanushikam@uom.lk

Abstract. Internet of things (IoT) is a new trend in today's technical world. Improvement of IoT devices makes life easy for human being since it reduces the complexity in their lives through converting impossible things to possible things. However, the exponential growth of IoT devices generates large amount of information by requiring an efficient technique to handle the huge amount of data in IoT. Thus, data mining has become the best solution to extract the huge information in a large data warehouse. With this background, this paper provides a comprehensive literature review on different data mining techniques, algorithms and models which can be used in IoT with their pros and cons. Further, this paper investigates most popular applications in IoT which use data mining techniques. Finally, this review is concluded by addressing some challenges in data mining for IoT in order to open the path for future researches in this field.

Keywords: Data mining, data mining algorithm, data mining model, data mining technique, Internet of Things

1 Introduction

In this era of technology, every field in the world is being developed continuously with new technologies. Internet of things or IoT plays a momentous role in this developing process since today's most technologies focus on the digitalization. IoT makes human life easy in the present than in the past by using various advanced computing technologies [1]. In recent years, IoT showed the great impact for the significant improvement of many of the fields like healthcare, agriculture, transport etc. IoT always deals with huge amount of data generated by these fields in order to develop smart applications. Therefore handling data in IoT is most important task which should be done in an effective way.

Handling large amount of data can be done via data mining which is a very popular technique in the field of data science. Since the sensors which are included in IoT devices collect and store data by creating a huge data base, when useful information is needed from this data base, data mining comes forward for fulfilling that duty in an efficient way [2]. Data mining is the process of extracting information from the huge amount of data stored in repositories or data warehouse or database [3]. Data mining concept has an important relationship with IoT because IoT cannot handle large data input properly without using data mining techniques. Data mining is a wide area that intersects with statistics, machine learning, artificial intelligence and database.

1.1 Data Mining

Data mining is the process of extracting the useful data, pattern and information by using different types of techniques [4]. In here generally, hidden information of large amount of data is found and expanded by using the best techniques. There are many terms in the field which can be used to interpret data mining. They are, namely, knowledge mining from databases, knowledge extraction, data analysis and data archeology. Traditional data mining methods were very expensive. Those methods were executed manually and therefore, processes were very slow. In contrast to the past, data mining techniques which are used in the present is very fast, accurate and automated [4].

Data mining process consists of three main stages as data preparation, data mining and data presentation. Data preparation is the main stage in the process. In this stage, data are prepared using three sub-stages such as, integration, extraction and preprocess. When gathering data using multiple sources, data may have different types. Therefore, these various types of sources are combined in a common source under the integration [4]. Data extraction is next sub-stage. There, some parts of the data are expanded in to data mining system [5]. Under data preparation, last sub-stage is data preprocess. Preprocess means preparing data in order to facilitate to next main stage. That is data mining. In this stage, appropriate algorithms or techniques are applied to identify the important pattern and information of the data which are available in the data stored or data set or data warehouse [3, 6]. Last stage of the data mining process is data presentation. At this stage, visualizing and representation of the information extracted from the data are carried out [3, 6].

1.2 Internet of Things (IoT)

IoT can be interpreted as a collection of connecting computing devices in order to make life easy for human through internet [6]. IoT is called as the next generation of the Internet. It contains large numbers of nodes which represent the various objects [7]. Mainly, IoT has an aim of building a superior network technology which automatically catches the requirements of users and will operate in that point of view [7]. IoT converts the complexity to simple things automatically. IoT devices are used by human for their day today life activities. There are sensors, actuator, internet enabled mobile devices, cars, house hold appliance, wearable devices, etc. [8, 9]. All of these devices have unique identifiers like IP address [9].

2 Data Mining for Internet of Things

This section focuses on describing the relationship between data mining and IoT. Different types of functionality in data mining which are used for IoT have been discussed here. Internet of Things (IoT) generates huge data with different types by using many devices. Therefore, large amount of data is needed to be expanded for different purposes. In that case, data mining is required in order to expand the useful information which generated by IoT as a more efficient and useful technique [9]. It reduces the complexity in the system for solving problems. Figure 1 shows how data mining model can be used in IoT when handling data in IoT.

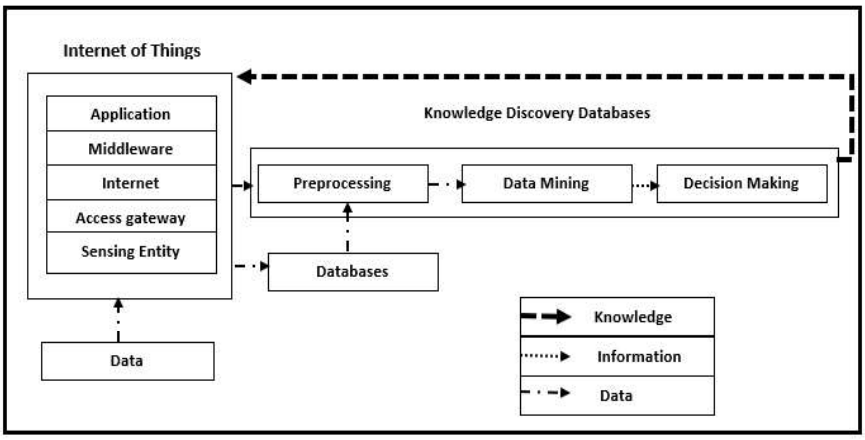

Fig.1 Internet of things with data mining model [10].

According to the figure 1, as a first step, data are gathered using IoT and it subjects to data mining process. Data mining process generates information from data. Then, it is sent again to IoT as knowledge. To increase the high performance of the data mining model in IoT, three keys are considered as objective, data and mining algorithm.

Objective is to identify the limitations, assumptions, measurement and issues of the problem. When using these details, it can be very easy to understand the objects of the problem [1]. In data mining, data play the most important role. Data have some characteristics such as size, distribution and representation [1]. When identifying those characteristics of data, it can be easy to determine the algorithm clearly. This has a significant impact on increasing performance of the system. There are many data mining techniques which use these algorithms.

2.1 Data Mining Algorithms for IoT

There are top mining algorithms which can be used in IoT. Those are C4.5 algorithm, K-means algorithm, Apriori algorithm, KNN algorithm, EM algorithm, Support vector machine algorithm and Page rank algorithm.

2.1.1. C4.5 Algorithm

C4.5 algorithm is used to build a form of decision tree. Therefore, C4.5 algorithm represents classification technique. Thus, this algorithm belongs to supervised learning method. As an example, we can consider set of patients. Each patient has age, blood pressure, pulse, family history etc. like attributes. If we want to select cancer patients, it will fall into whether patient has cancer. C4.5 categorizes class for each patient [11].

2.1.2. K-means Algorithm

K-means algorithm is used for clustering. Clustering is creating groups having similar members. K-means creates k number of such groups from set of objects. Therefore, this algorithm belongs to unsupervised learning method. However, this algorithm is sensitivity to the outliers. It can be considered as a weak point in this algorithm [10, 11].

2.1.3. Apriori Algorithm

Apriori algorithm is used to execute association rule and this applies for large number of transactions. This algorithm is understandable well and also implementation is very easy. This algorithm can be defined as unsupervised learning method. As an example, this algorithm can be used for supermarket transaction. It contains customer transaction and different grocery items which have been included in the spreadsheet [11].

2.1.4. KNN Algorithm

K- Nearest Neighbors (KNN) algorithm is used for classification but it is differing from the classify because it is lazy learner. Lazy learner means that it does not do much during the training process other than storing the training data. Therefore, KNN does not build classification model, but stores the labeled training data. KNN is supervised learning because it creates a labeled training data set [10, 11].

2.1.5. EM Algorithm

Expectation-Maximization (EM) algorithm uses for clustering. EM generates the best model that assigns class

labels to data point. It is unsupervised learning method because it does not provide labeled class information. EM specifies for clustering and for generating a model with parameter [11].

2.1.6. Support Vector Machines Algorithm

Support Vector Machine (SVM) algorithm is supervised learning method and belongs to association rule in data mining. It is used for regression analysis and classification [10, 11].

2.1.7. Page Rank Algorithm

Page Rank (PR) algorithm is designed for some objects which are linked within a network object. Google search engine is main example for page rank algorithm as well as other worldwide web (WWW) use this algorithm. This is unsupervised learning approach because it is used to discover the importance or relevance of the web page [11].

2.1.8. ANN Algorithm

ANN algorithm or Artificial Neural Network is a model of machine learning that is used to learn human brain. ANN has basically 3 layers. There are, an input layer, a hidden layer and an output layer [12]. However, the "signal" passes at a connection is a real number and output of every neuron calculated by using non-liner function of the sum of its inputs. Generally, above mention neuron aggregates with above layers. ANN algorithm plays important roles in the different fields such as computer vision, speech recognition, machine translation, social network filtering and medical diagnosis. ANN algorithm has some components. There are neurons, propagation function, connection and weight [13]. Under ANN, deep neural network (DNN) plays an important role since it has a large number of layers. There are input layer, output layer and many hidden layers in between. There are specific type of sorting method and ordering function in above mention layers. However, there is some trouble in this algorithm. It means, algorithm always works with unlabeled data or unstructured data. Generally, this algorithm combines with machine learning. Simply, it tries to classify and order information that go beyond simple input/output protocols [14].

2.2 Data Mining Techniques for IoT

IoT generates huge amount of data which have very complex nature. Therefore, advanced techniques or functionalities are required to handle these data. Since, the data which are generated by IoT has different types, finding a suitable method for analysis and extracting data is needed. There are several data mining techniques which can be used for this task. There are classification, clustering, association rule, outlier, and time series. Further, data mining techniques can be categorized as descriptive and predictive [15]. Clustering, association rule and outlier can be considered as descriptive data mining techniques while classification and time series go under predictive data mining techniques. Descriptive model defines overall probability distribution of the data, partition the whole data into groups and describes the relationship between the variables [15]. Predictive module permits the value of one attribute / variable is to be predicted from the known values of another attribute / module [15].

2.2.1 Classification

Classification is used to classify both labeled and unlabeled pattern [1]. Part of unlabeled data which has some information, class or tag is considered as labeled data [1]. Classification algorithm first builds the model by using training data and as a second step, it uses the test data to check the accuracy of the model. After certifying the accuracy of the model, it can be used to classify the unlabeled data [15]. However, training data should have a set of attributes with known class labels [15]. Further, classification can be defined as supervise learning method because classification techniques represent data item as labeled classes and also can be defined as an important algorithm in machine learning which can be used for forecasting [15]. Classification can be used to find a category of the particular item in a dataset in an easy way [7]. Figure 2 shows the classification process in data mining.

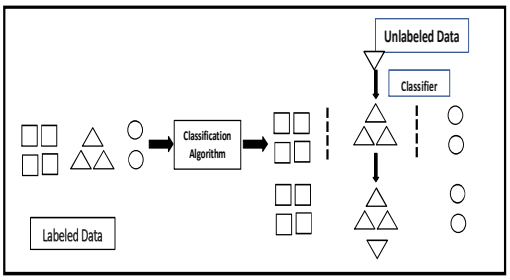

Fig.2 Classification process in data mining [1].

Before classifying the unlabeled data using classification algorithm, labeled data input creates a set of classifiers. Then, unlabeled data pattern classifies the set of classifiers [15]. Therefore, classification algorithm can partition all input patterns very easily.

2.2.2 Clustering

Clustering is a member of descriptive group. Clustering is used to classify unlabeled data patterns [1]. Therefore, clustering technique can be defined as unsupervised learning method. Figure 3 indicates clustering process in data mining. In here, input data patterns are partitioned in to three kind of groups as labeled data pattern according to their similarities [15].

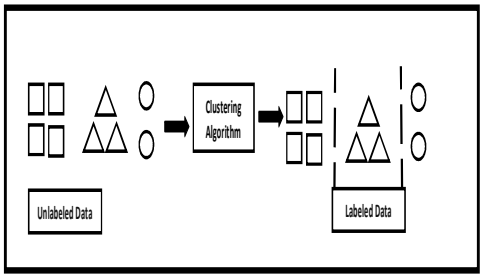

Fig.3 Clustering process in data mining [1]

2.2.3 Association Rule

Association rule is a member of descriptive family. This is used to find a specific event that does not occur in a particular order [1, 7]. It is known as frequent pattern mining process [1]. As the name suggested, frequently happened pattern in data stored or data base or data warehouse is found by this rule. Figure 4 shows the association rule process in data mining.

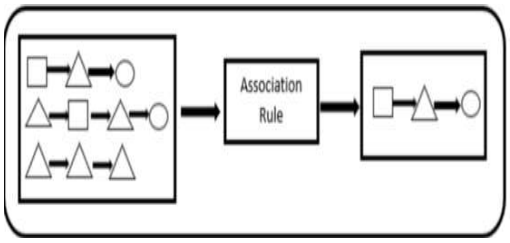

Fig.4 Association rule in data mining [1]

2.2.4 Outlier Study

Outlier can be found by calculating distance among objects in the data. Therefore, an algorithm used for outlier study is based on the distance [3]. Outlier is an observation that deviates from an overall pattern of the sample data [11]. There are two types of outlier as univariate outlier and multivariate outlier. Univariate outlier can be found when looking at a distribution of values in one space while multivariate outlier can be found by looking at distributions in multi dimensional spaces [16]. Specially identifying multivariate outlier is difficult without using a model. That is the reason to train an algorithm in data mining for outlier study [16].

2.2.5. Time Series

Time series is a member of predictive groups. It predicts a rough calculation of future data by analyzing collected data [7]. Generally, time series has three components. There are, representation, similarity measure, and indexing [5]. Representation is mainly used to reduce the dimension of the time series. Similarity measure is an approximated measure which calculates the similarity between time series [17]. Indexing part is closely connected with representation and similarity measure [5].

2.3 Data Mining Models for IoT

Data mining process includes four steps as data collection, data preparation, data mining and data presentation. These four steps of data mining have been divided in to four categories [18]. First one is IoT with knowledge discovery model. This is used to show how the raw data have been converted into useful information [19]. Second one is multilayer data mining model. It does data collection, data processing and data mining services [19]. Third model is distributed data mining model. This model is used to service and analyze the data stored in the distributed locations [18]. Last model is grid-based model. Grid based model is applied for the large amount of data [19].

2.3.1. IoT with Knowledge Discovery Model

IoT with knowledge discovery model is similar to data mining knowledge discovery model. IoT devices produce huge amount of data store. When applying knowledge discovery model in IoT, it extracts useful information which is used to generate knowledge. This model has four sub steps as follows [18, 19];

1. IoT database – includes sensor data, application data, data from other resources.
2. Data mining step – extracts useful information from the large data base.
3. Data processing step –converts information to knowledge for decision making.
4. Decision making step – performs future forecasting in order to get decision.

This model is very important for a system in order to improve the performance of the system and service quality [19].

2.3.2. Multilayer Data Mining Model

According to data mining model architecture, next mining model is multilayer data mining model. This model divides into 4 subparts. These subparts can be defined as [2];

1. Data collection layer - consists of IoT devices such as radio frequency identification (RFID) readers, sinks etc. and different type smart object's data such as global position system (GPS) data, quality data, RFID stream data, satellite information, positional data etc.
2. Data management layer - manages distributed or centralized data base or data ware house [18].
3. Event processing layer - distributes integrated data and provides high quality of data [18]. Event filtering is most important task in this layer.
4. Data mining service layer - provides an efficient solution to the problem by separating parts of the complex problem. Further, this layer facilitates data cleaning and merging [20].

2.3.3. Distributed Data Mining Model

Internet of things with discovery knowledge data mining model and multilayer data mining model are used basically to handle the local data [19]. However, distributed data mining model explains distributed infrastructure. Distributed architecture model always deals with the different nodes which are coming from multiple sites. Mainly, it contains two types of layers. There are global and local layers. Data mining process is performed in local model and global model and helps to visible this update to all nodes [19]. Global control layer is the core of

the main system. The tasks of nodes are, selecting the data mining algorithm, data sets for mining and navigating sub nodes in the data set [2]. After sub nodes process all of the data, the data are saved in local warehouse or database or repository. The local models are acquired by event filtering, complex event detection and data mining in local nodes. After demanding the local nodes as a requirement, local nodes are given into global node and these two types of nodes aggregate together. The whole process is manipulated by multi agent collaborative management mechanism [2].

2.3.4. Grid Based Data Mining Model

Grid based data mining model can be defined as a grid manner. Therefore, this model is used for multi model sensors data. The grid-based mining model facilitates parallel task execution or parallel data mining. Grid-based model provides a facility for the user to use computing resources of grid. It acts as power resources [2]. This mining model consists of five types of layers such as IoT resources layer, IoT service layer, grid middleware layer, grid mining layer and grid application layer [19].

3 IoT Applications which use Data Mining

There are so many applications which can be seen in the various industries and fields which use IoT devices and sensors which have been connected through the internet. Bank, healthcare, smart city, vehicle system, etc. are such fields [7]. In this section, most popular IoT applications which use data mining have been described such as, smart city application, smart home application, enterprise application, health care application, and E- commerce application.

3.1 Smart City Application

When using data mining for IoT in the smart city application, there are several systems in the field. There are traffic control, suburban electronic meters and pipeline leak detection.

Traffic control has become a huge problem in city life. When developing a system which aims at this kind of problem, information relevant to the traffic flow in the city can be gathered by using IoT devices such as GPS, vehicle system and smart phones. Information may include time of travelling in particular area, accident areas, construction areas, heavy traffic areas, etc. This problem can be defined as traffic blockage problem and can be solved by using classification algorithm. Classification algorithm can use the information generated by IoT and then can classify selected area. Then the system can provide the best solution for that particular traffic problem through the knowledge obtained via classification algorithm [7].

Sub urban electronic meters use smart devices to measure energy and hence, relevant energy consumption. As per the traditional method, people should obtain bill information from the meter reader. According to smart city concept, they obtain information through digitally such as through E- mail or through smart phones. Further energy consumption of daily or weekly or monthly can be predicted as a customer requirement by using time series algorithm [7].

Pipeline leaking is another problem in the metropolitan area since many constructions and also maintenance activities happen frequently in these areas. This problem can be solved very easily by using outlier series technique. Along with this system, use of sensors allows to detect spot leaks areas [7].

3.2 Smart Home Application

Concept behind the smart home application is controlling the processes automatically according to certain time period by using IoT devices [7]. Smart home application includes the central computer control system. Different types of electronic and electric devices can be connected to the control system. Then, all the processes happen in a time framework. There are many applications which can be identified under the smart home applications such as smart water application, environment based application, smart grid application and smart infrastructure application [7].

Mainly, classification mining technique and time series mining technique are used in smart home application. Data which processed by IoT devices in these applications are stored with respective to the related time period. Time series technique is used for this process. Finally, future situations will be predicted by applying regression [5].

3.3 Enterprise Application

If an organization or company wants to find the best solution to handle the business processes, enterprise application can be used. The enterprise application acts as a major control hub and handles all subparts of the business process. It receives large amount of data from the different sources and it will cover all of the functional areas in a business such as human resources, finance section, manufacturing section, distribution section, transportation, maintenance, sales and marketing section etc. [5]. Use of IoT and data mining to enterprise side provides different types of services. There are smart banking application, smart retail application, energy based application, smart manufacturing applications and smart agriculture applications [7].

3.4 Health Care Application

When considering health problems of human being, it cannot be handled manually, because they have millions of gens and all gens are composed to individual nucleotides [7]. So, data mining technology can be used to reduce the complexity of the situation. By using sequence pattern mining or association rule mining technique, gens can be analyzed and similarities can be searched. Thereafter, different diseases can be easily identified [5].

Under the health care application which is related to IoT and data mining, there are different types of applications. These applications are based on the various services in this field such as, emergency ambulance services, hospital management, emergency room, research, elder care, billing equipment, clinic based

services, surgery lab, diagnosis, etc. [7]. All the data stores by using intelligent system. Then, accurate information can be provided to patient. By using patient's report and other historical details, it can be very easy to predict the patient's diseases situation. In here, clustering data mining technique can be used [5].

3.5 E-commerce Application

E-commerce concept in data mining has become very popular in the recent years because it helps for mining hidden pattern in transactions [5]. These details help to plan and organize new markets in a very easy way. Massive amount of data in the data bases in this field contain customer data, product data, user's action log data etc. Association analysis technique and clustering mining technique are used generally in the field of e-commerce in order to motivate customers to purchase products or services which they missed [21]. E-commerce application contains several sub parts which uses IoT with data mining. There are customer profiling, recommendation system, web personalization, and customer behavior [21].

By considering all the applications mentioned above, table 1 shows the frequently used data mining techniques in these applications.

Table 1. Frequently used data mining techniques in IoT applications [5].

IoT Application	*Classification*	*Clustering*	*Association Rule*	*Time Series*	*Outlier Analysis*
Smart city	√	√	√	√	√
Smart home	√			√	
Enterprise	√	√	√		
Healthcare		√	√		
E-commerce		√	√		

According to the figures in table1, it is shown that clustering and association rules are the most used data mining techniques in this type of IoT applications. Since smart city contains lots of sub-systems, all of the data mining techniques should be used to strengthen this kind of IoT application.

4 Pros and Cons Analysis for Data Mining Techniques and Algorithms

4.1 Comparison of Data Mining Techniques

Classification, clustering, association rule, outlier study and time series are the top data mining techniques which have been discussed in this review. These techniques are further described in this section according to their pros and cons.

When considering classification technique, it labels the data. Therefore, it is very easy to do the analysis. Classification method always processes parallel to their hardware [22]. Therefore, it is very easy to handle. However, it is a time-consuming technique. Clustering technique is used to classify unlabeled data. Main advantage of clustering is recovering failures automatically. Other than that, clustering measures all of the hierarchical levels of data [23]. However, there are some disadvantages also. Mainly, it is very complex process [24]. Then, it will be difficult to find correct number of clusters that the user wants. Sometimes, information will be lost due to heavy outlier [25]. Association rules very different than other data mining techniques. It can mine data quickly without wasting time or any other hardworking. All the processes have flexibility. Other thing is that, it can produce many rules by using lots of information as user wishes. One disadvantage in here is slight deviation from the reality. It happens due to the assumptions used when applying this rule [26]. Outlier study plays an important role in data mining because it finds irregular patterns in huge datasets which have abnormal behavior [27]. There are various outlier detection approaches in data mining such as clustering based approach, classification approach, statistical approach and proximity based approach [28]. Under these approaches, there are different types of methods to perform outlier analysis. Advantages and disadvantages of using outlier analysis vary according to those methods. Some methods are very complex while some have less complexity. Some methods perform in a very effective way on high dimensional data while others perform poorly. There are less efficient techniques as well as high efficient techniques [28]. Time series techniques are basically used for forecast purposes. So, the main advantage of the technique is that, it can predict future approaches by using past behaviors of the situation. Time series helps to compare the performances of various types of series at the same time. Main problem with time series is that, it is not always ideal. So, some conclusion may be uncertain. If the data are not changing continuously, it may hard to predict the future directions [29].

4.2 Comparison of Data Mining Algorithms

Top data mining algorithms which have been described in the above are described in this section with respect to their pros and cons.

C4.5 algorithm is very easy to understand than others because all data are classified and represented by using decision tree. The provided information is very accurate and it can be used for future directions. C4.5 algorithm also has some limitations. For decision making purposes, it will test one attribute at a time. Thus, this is time wasting algorithm when process handles the huge scale of data set [22]. K-means algorithm is based on the clustering techniques. When there is a huge dataset, k-means algorithm can be used because of its high performance. Further, it is very easy to implement. However, it has some difficulty when finding the optimal solution, but if there is a largest amount of data collection, without any matter, it

can be measured in an efficient way [30]. Apriori algorithm is applied in association rule. This algorithm is very easy to implement. One disadvantage of the algorithm is needy of several iterations when data are processed [30]. KNN algorithm is also based on classification technique. This algorithm always deals with the non- parametric architecture. This is simple and powerful algorithm than others. It does not require training time, but estimation is slow and memory is accelerated [31]. EM algorithm is used in clustering algorithm. This is very easier algorithm than others and also this has less dependency on initial parameters but has more dependency on all the sequences. For searching initial parameter spaces, it will take very less time [32]. Support vector machine algorithm can be used easily to partition data and that is very accurate method for prediction. However, it consumes time due to space and size allocated for time training in testing. Since this algorithm can identify liner transformation, it will make extracting data in a very easy way [22].

5 Issues in Data Mining When Applying for IoT

Data mining for Internet of thing is a huge process in the field of data science. Therefore it contains some issues. Identifying these challenges will make directions for future researches. With that aim, most significant issues with data mining are listed below;

5.1 Accessing Large Storage of the Data from Various Data Storage Location

IoT devices gather data using different types of equipments. Then, these data are stored in the various locations. When it is processing, it faces massive cases because of the inability of IoT devises to detect data properly [5]. Further, data mining in IoT deals with different types of data such as mixture data, incomplete data, noisy data etc. In that case, issue will happen with traditional algorithms since these algorithms are not suitable for these situations [5].

5.2 Privacy and Security

Privacy and security are another main problem in data mining with IoT. When user gathers information, it gives serious problem to the user unknowingly regarding privacy and security [33]. As an example, we search many things using Google or whatever search engine. Then, they save history and send various type of advertisement to the user without user's permission. That is because it spreads data to another related organization or company [1].

5.3 Infrastructure perspective.

Data mining is managing large data sets and also it is adding continuously data to the data base. Then traditional software tools cannot handle this situation properly [34]. In here, IoT gives law power and result is high, but role of mining algorithm acts as opposite to IoT. According to above situation, it creates infrastructure perspective.

5.4 Mining is uncertain and incomplete

In data mining, it uses various types of data such as mixture data, incomplete data, noisy data, etc. Since data mining involves with large amount of data and it may not include always correct data or reliable data. Then, the results obtained through mining these less reliable data will not be completed and also predictions will not be accurate [5].

6 Conclusions

IoT and data mining are very important techniques in the field of new technology. Gathering data by IoT and extraction are needed to do in the best way since IoT always deals with the huge amount of data. Thus, it is essential to use data mining in IoT. Specially, several data mining techniques can be applied to extract these data. Even though, classification technique is referred as a best technique in the literature, according to this comprehensive literature review, suitability of the data mining technique for the given situation will depend on the particular scenario in which they are used. The main reason for that is having different characteristics for all the data mining techniques. According to the pros and cons analysis, most suitable algorithm also will be changed based on the scenario because these advantages and disadvantages depend on the situation they are used. Finally, main issues in data mining have been summarized and described in this paper in order to open path for future research directions.

References

1. Prajapati, P., Vasa, J. and Patel, J. (2017). A Research Direction on Data Mining with IOT.
2. Bin, S., Yuan, L. and Xiaoyi, W. Research on Data Mining Models for the Internet of Things. 978- 1-4244-5555-3/10 (2010).
3. B O, V. and Mungara, J. (2017). Data Mining Algorithms for Internet of Things: Technical Review. *IRIT*, Vol.5(Special Issue 5).
4. Gera, M. and Goel, S. (2015). Data Mining - Techniques, Methods and Algorithms: A Review on Tools and their Validity. International Journal of Computer Applications, 113(18), pp.22-29.
5. Chen, F., Deng, P., Wan, J., Zhang, D. Vasilakos, A. and Rong, X. (2015). Data Mining for the Internet of Things: Literature Review and Challenges. International Journal of Distributed Sensor Networks, 11(8), p.431047.
6. Mohmmad, S. and Sunil, G. (2017). A Survey on New Approaches of Internet of Things Data Mining. International Journal of Advanced Research in Computer Science, 8(8), pp.666-673.
7. Krishna, G. (2017). Data Mining Processes, Applications and Challenges for IoT. International Journal of Trend in Research and Development, Volume 4 (3) (ISSN: 2394-9333).
8. Wlodarczak, P., Mustafa, A. and Soar, J. (2017). Data mining in IoT: data analysis for a new paradigm on the internet.
9. Gupta, P. and Gupta, R. (2017). Data Mining Framework for IoT Applications. International Journal of Computer Applications, 174(2), pp.4-7.

10. Yedukondalu, B. and Raju, D. A. (2017). New approaches of Data Mining for the Internet of Things with systems: Literature Review and Compressive. International Research Journal of Engineering and Technology, Volume: 04(Issue: 08).
11. Kdnuggets.com. (2015). Top 10 Data Mining Algorithms, Explained. [online] Available at: https://www.kdnuggets.com/ [Accessed 15 Sep. 2019].
12. Inthachot M, Boonjing V and Intakosum S. (2016) Artificial Neural Network and Genetic Algorithm Hybrid Intelligence for Predicting Thai Stock Price Index Trend,Computational Intelligence and Neuroscience Volume 2016, Article ID 3045254, 8 pages.
13. Wikipedia,Atificial neural network[online].Availbleat:https://en.wikipedia.org/wiki/Artificial_neural_network.
14. Wikipedia.Deep Learning[online].Availableat:https://en.wikipedia.org/wiki/Deep_learning
15. Dogra, A. and Wala, T. (2015). A Review Paper on Data Mining Techniques and Algorithms. International Journal of Advanced Research in Computer Engineering Technology, Volume 4(Issue 5).
16. Santoyo, S. (2017). A Brief Overview of Outlier Detection Techniques. [online] Towards Data Science. Available at: https://towardsdatascience.com
17. Yoshida S. and Chakraborty B. (2017), A Comparative Study of Similarity Measures for Time Series Classification. In: Otake M., Kurahashi S., Ota Y., Satoh K., Bekki D. (eds) New Frontiers in Artificial Intelligence. JSAI-isAI 2015. Lecture Notes in Computer Science, vol 10091. Springer, Cham. [online] Available at: https://link.springer.com/chapter/10.1007/978-3-319-50953-2_27
18. Selvi, S. (2018). Difficulties and Data Mining Model for Internet of Things (IoT). International Journal of Engineering Science Invention (IJESI*)*, (2319 – 6726), pp.44-48.
19. Joshi, D., Kumari, C. and Srivastava, A. (2016). Challenges and Data Mining Model for IoT. International Journal of Engineering Applied Sciences and Technology, Vol. 1(Issue 3), pp.36-41.
20. Savaliya, A., Bhatia, A. and Bhatia, J. (2018). Application of Data Mining Techniques in IoT: A Short Review. National Conference on Advanced Research Trends in Information and Computing Technologies, Volume 4(Issue 2).
21. Hongjiu, G. (2013). Data Mining in the Application of E-Commerce Website. Intelligence Computation and Evolutionary Computation*, AISC 180*, pp.493-497.
22. O. Yamini and Prof. S. Ramakrishna (2015). A Study on Advantages of Data Mining Classification Techniques. International Journal of Engineering Research and technology, V4(09).
23. Docs.oracle.com. (n.d.). Advantages and Disadvantages of Clustering. [online] Available at: https://docs.oracle.com [Accessed 15 Sep. 2019].
24. Bhagat, A., Kshirsagar, N., Khodke, P., Dongre, K. and Ali, S. (2016). Penalty parameter selection for hierarchical data stream clustering. pp.24-31.
25. Anon, (n.d.). What are the disadvantage of clustering in data mining? [online] Available at: https://www.quora.com/
26. Slideplayer.com. (2019). 1 CISC 4631 Data Mining Lecture 09*:* Association Rule Mining Theses slides are based on the slides by Tan, Steinbach and Kumar (textbook authors) [online] Available at: https://slideplayer.com/slide/7494454/
27. Zhang, J. (2013). Advancements of Outlier Detection: A Survey. ICST Transactions on Scalable Information Systems, Volume 13, Issue 01-03. [online] Available at: https://core.ac.uk/download/pdf/11049855.pdf
28. Modi, K. and Oza, B. (2016). Outlier Analysis Approaches in Data Mining. International Journal Of Innovative Research In Technology, Volume 3, Issue 7. [online] Available at: https://pdfs.semanticscholar.org/526d/2be3a950b9d60edf87d9cbb6ffcc8f5d4a08.pdf
29. Thephysicspi.blogspot.com. (n.d.). Advantages and Disadvantages of time series. [online] Available at: https://thephysicspi.blogspot.com/2019/02/advantages-and-disadvantages-of-time.html [Accessed 15 Sep. 2019].
30. Mehta (n.d.). K means clustering algorithm. [online] Slideshare.net. Available at: https://www.slideshare.net/DarshakMehta6/k-means-clustering-algorithm-63472248
31. Slideplayer.com. (n.d.). K Nearest Neighbors Classifier & Decision Trees - ppt video online download. [online] Available at: https://slideplayer.com/slide/7863810/
32. Slideplayer.com. (n.d.). Expectation Maximization and Gibbs Sampling – Algorithms for Computational Biology Lecture 1- Introduction Lecture 2- Hashing and BLAST Lecture ppt download. [online] Available at: https://slideplayer.com/slide/7671025
33. Wideskills.com. (n.d.). wideskills.com | Challenges in Data Mining. [online] Available at: https://www.wideskills.com
34. Jaseena, K. and David, J. (2014). Issues, Challenges, And Solutions*:* Big Data Mining.