International Journal of Control Theory and Applications

ISSN : 0974–5572

Volume 9 • Number 42 • 2016

Review on Recent Research in Data Mining based IoT

Supriya Sarkar[a] and Chandra Mohan B[b]

[a]*Research Scholar, School of Computing Science and Engineering, VIT University, Vellore, India.*

E-mail: supriya.sarkar@rediffmail.com

[b]*Associate Professor, School of Computing Science and Engineering, VIT University, Vellore, India.*

E-mail: dr.abc@outlook.com

Abstract: The growth of IoT in the real life is increasing very rapidly in areas like smart cities, IT, manufacturing industries and health monitoring. While data are sensing from IoT devices which are uncertain and unstructured format. Data is generated very rapidly and very huge data needs to be processed everyday to derive some output by using various data mining technique. Data mining technique is one of the important steps in Knowledge Discovery from Data (KDD). Traditionally, knowledge is extracted manually and so it was extremely costly, time consuming and error in data occurred. At present, certain Data Mining(DM) techniques like clustering, association rule, outlier detection are revolving in the researchers and industry community. Applying the DM technique along with challenges like data accuracy, uncertainty, massive data, overhead of storage and communication on IoT, lot of issues will occured. Knowledge discovered as pattern value from data mining technique, processing, transformation, cleaning are applied in huge database which are used by IoT sensor device for storing data. This paper reviewed KDD with various applications like neural network, air pollution and spatial data mining. Also this paper focuses data mining with IoT, its challenges and need of optimization in different real time application.

Keywords: *IoT, KDD, Data Mining, Hadoop.*

1. INTRODUCTION

It is expected to grow exponentially from billion to trillion devices connected with each other by Internet of things in the year 2020. The internet has changed drastically and moving towards a new era from professional to personal life. It has a smart device like RFID which sense intelligently, tracking location, monitoring as a dynamic network. The RFID is a device which communicates with a tag by unique address identification scheme suggested by [1]. Not only machine to machine or device to smart device communicated, but also machine to human or vehicle to vehicle communication is made possible in Internet of Things[2].

There are three visions in the Internet of Things which are internet oriented vision, semantic oriented vision and things oriented vision works, defined by [3]. Various applications are mentioned in different domains such as smart environment domain, transport domain, healthcare domain and personal domain. IPv4 is a 32 bit address which majorly used for computer network and IPv6 is a 128 bit address used for recent mobile computing and the same is used in IoT as it has huge addressing space.

Many growing industries like IBM, which are working with IoT technologies for analytic or managing their huge real time data. By using IoT, in the next future, smart world with smart cities, smart industries with smart devices, are expected. These IoT's intelligently handle huge data, uses low power and lower internet utilization cost. There are some challenging issues in IoT such as Data Management, Unique Addressing, Security and Privacy, Power and energy storage and Network Management by [4].

2. RELATED WORK

2.1. Knowledge Discovery

Huge amount of abundance data are available in the network and which is unnoticed. It needs some data analysis tool for analyzing that data, because a real time data available in database are in poor situation. Large amount of data are generated from sensor device and available in database. Traditionally, knowledge is extracted manually and so it was extremely costly, time consuming and error in data occurred and this was proposed by [5]. Although the rich data are stored and extracted valuable knowledge data for fast growing and trying to access extracted knowledgeable data from huge database by the knowledge discovery from data steps *i.e*. knowledge mining.

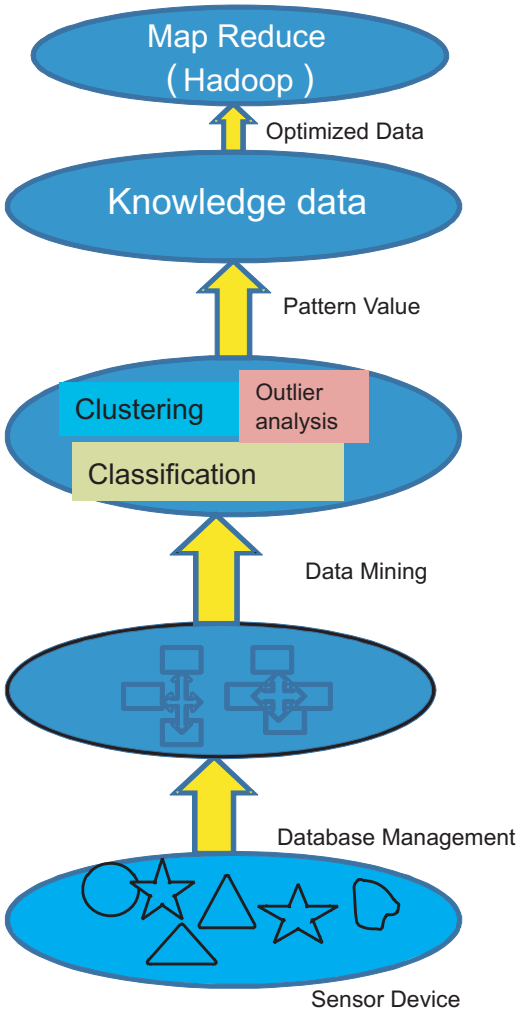

Figure 1: Flow of process in the IoT based knowledge discovery

KDD process having number of steps such as Data cleaning, Data Integration, Data Selection, Data Transformation, Data Mining, Pattern evaluation and Knowledge representation. In Fig. 1. described step by step process, by sensor device data are sensing then store into database which are unstructured style that managed by database management system. For processing the unstructured or uncertain data, data mining techniques are applied. The data mining techniques are like clustering, association rule, outlier detection, classification etc. For generating knowledgeable data optimization tool is used like MapReduce which is used to produce optimal pattern information [6][7].

Different fields are mentioned in KDD which are applied with IoT like neural network, spatial data mining, air pollution, health monitoring, IT industry etc. as shown in Fig. 2. Each application having its own functionality to discover knowledge from real time huge database.

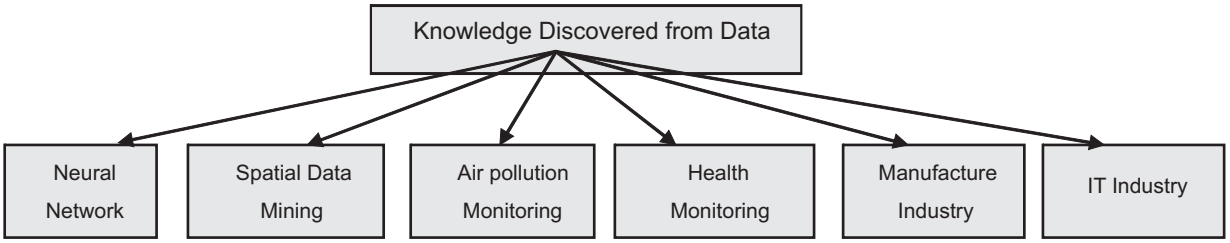

Figure 2: KDD in various field of IoT

2.2. Neural Network

[7] suggested a system for retrieving new knowledge information according to prior knowledge. As all the information are not useful which mined by data mining technique. So, for improving the access frequency of relevant data from database, author proposed cognitive map. By using KBNMiner in cognitive map, knowledge rate is predicted and system having functionality like prior knowledge keyword which used to predict relevant keyword and check with the keyword set. After that, prediction of positive interest is made by using this Ekt = Pkt/ (Pkt + Nkt) having 0.5 rate which was predicted as highest interest over negative interest. While comparing previous system with this technique, the value 0.527 is obtained as average error which was maximum error in knowledge. In this proposed system author has compared two vector matrix for predicting error. One matrix having knowledge without neural network N1 and other matrix having knowledge with neural network N2. So the maximum error free is achieved in second matrix N2.

2.3. Spatial Data Mining

[8] Proposed spatial data mining which plays vital role for smart earth for low population, low environment impact and low power consumption. By proposed architecture data retrieve from earth in the form of map, pixel, image, text, audio then through acquisition and analysis equipment stored into spatial database. Then the refined unknown data from database is converted into information. Then, collected information is processed for knowledge data by data mining technique like crisp rule and which extends into intelligent pattern. This proposed technology is applied for improving the relationship between nature and human.

2.4. Air pollution Monitoring

[9] Proposed that IoT introduced the large number of heterogeneous data managed for different application like environment monitoring, air pollution by using service controlled networking tool. In proposed system SCN layer describes description between services and application. When user request any application, it should be decided about what service provided to map by message exchange (*h_mep*), service discovery(*h_dscv*) and service status function(*h_stat*).Whether requested service is available or usable, accordingly it is formalized

using So(*h_mep, h_stat, h_dsev, q*) = {*m*/*m* = *h_mep*(S1….S*n*), Si ∈ S}, service discovery function S = *h_dscv*(*q*). Selecting the services satisfying the status *h_stat*(*s*) = TRUE. Then finally messages are exchanged between services and application. STICKER presented knowledge discovery and information enable the physical sensor data (PM 2.5) with social sensor data and generate relationship for forming systematic solution for managing sensor big data and extract valuable information. According to case study by author, in 2013 at Japan heavy air pollution occured due to traffic jam. So the solution is to identify by defining different colour and cross checking each other according to sensor and service.

2.5. Health Monitoring

[10] Proposed for health lifestyle, for monitoring patient, in this model knowledge acquisition is used for identifying the device, machine and all the people connect with that device. The data model used to provide the facilities of knowledge mining from ubiquitous computing environment. After analysis by data mining technique simplified knowledge is derived and it makes as a scalable solution. The data model located in patient home and another model is located to medical purpose at medical centre. Prior system did not support for function like medical error, remote device managing. So by using YOAPY tool in the data model, it reduces limitation of overhead in medical error. It is aggregated domain knowledge from raw data and compression technique applied on them. While using wearable device in patient body for measuring ECG, time series is used in raw data. Then the time series interval A is divided into sub signal B. By this proposed tool YOAPY get 300 samples/sec and which is more accurate data according to previous model.

2.6. Manufacturing Industry

[11] author proposed for manufacturing process in industry due to impact of environment, knowledge discovery is used for energy efficient and quality issue at management and supply chain. Not all the data is useful for data mining, so KDD is applied to find valuable pattern from dataset. In manufacturing industry, devices are operated by RFID, and then data is gathered into database and check whether it is field related data or decision making data. After storing data, KDD method is applied on it. Then get the knowledge able planning of manufacturing process and move to micro grid power planning system. Hence, optimization process is generated from services and this optimization can be applied in real time application. The impact of environment is less as well as energy management in power supply chain also is efficient by using Micro grid planning. In previous method, it does not reply unexpected conditions which occur in industry, but by using micro grid technique replying quickly for unexpected condition is observed. So KDD technique is useful to make significant contribution, supervision and planning for power system to make easier life for power companies.

2.7. IT Industry

[12] Suggested a method called context aware sensor middleware (CASCOM), to provide semantic knowledge with accuracy, reliability, cost over raw sensor data. Design the IoT middleware for non IT and IT expert. It support IoT algorithm which is very difficult to understand by Non IT expert. CASCOM is semantic model, which means if new sensor added in model then also it will not affect on algorithm. Proposed model work like device sensing at physical layer, then context or semantic discovered from semantic layer, later processing the semantic knowledge and retrieve context knowledge by primary context processor. After data dispatcher or publisher send the data in the form of knowledge, send to database or analytic multimodal or cloud database.

Finally non IT expert retrieve knowledge from database. According to result by proposed model amount of time required is less than 200 Seconds. It takes 1GB data to model and store 100,000 IoT resource descriptions. As without CASCOM model user do not know how to retrieve data from database? Which sensor used for retrieves data and don't know how to configure? This problem can be overcome by IoT middleware. By using this model select question like temperature sensor detect by air stress and request the possible answer by clicking the search task. Searching query can return result in less than 1.53 sec. It is feasible to use real world application.

2.7.1. Data Mining

Data mining is the concept refers to mine an interesting pattern and discover knowledge from large amount of data. There are diversified varieties of data which is applied for data mining like data stream, ordered sequence, graph, spatial data, text data and multimedia data. Those data which are discovered might be structured data or unstructured data. For processing discovered data, there are different techniques like classification, clustering, outlier analysis.

[13] Discussed about unpredictable item from raw database. As Association rule won't support for finding minimum support based on Apriori algorithm, the rare association rule will predict the support properly and get the valuable knowledge. Accuracy, qualities are the main issues which are maintained by PCKA method which configure between relevant and non relevant data. Hence, many method like K-mean, Fuzzy c-mean clustering Technique are used to find relevant and non-relevant data, which was discussed by [14].

2.7.2. Data Mining with IoT

IoT is a concept where large number of device is interconnected with each other. The device as object or things depends on purpose of application. It samples the data continuously from different location where objects are identified uniquely based on RFID tag. There are many applications where sensing devices are used in different arena such as building, vehicle, health, automation. So that sensing raw data are in massive and this is generated from different location and this was discussed by [15].

That massive data are produced automatically from real time application so need to mine it and get the valuable knowledge by different technique of data mining algorithm, it may be clustering or classification, etc depend on the application. Though smart objects are used to sense data for different service like manufacture industries, environment monitoring, and health monitoring, event detection. This was discussed by [16].

IOT is expected to generate large amount of data from diverse locations that is aggregated at very high velocity. It demands better methods for indexing, storing and processing such data which in turn requires developing a technique that converts this data into a knowledge based defined by [17].

Number of layers mentioned for extracting knowledge by [18] like sensing layer for sense the data, data management layer for managing the data and data mining layer for valuable knowledge data. There are some characteristic of big data in IoT like Multisource, heterogeneity, huge scale, uncertainty.

3. RESEARCH ISSUES IN DATA MINING WITH IOT

Many challenging issues are discussed by researcher in data mining with IoT like data quality, uncertainity of data, inaccuracy of data, frequently updated time stamped structure and massive real time data observed by author [19].

3.1. Data quality

[20] Represented outlier detection technique used for error detection, using data mining technique to make smart decision and provides services for overcoming inaccurate and invisible information. Data cleaning and outlier detection are used for purifying data and provide quality data. Outlier detection improves the quality of data as data are in unorganized manner in database. Consider 0 & 1 outlier degree. 0 outlier means normal measurement and if more than 0 means anomalies are present and need to remove it.

Through one of the example explained by author that e-health system having outlier detection detect anomalies properly instead of other technique because there is no chance to find error in patient treatment otherwise small error spoil life. So, data cleaning technique can be used with outlier detection for detecting error with 100% accuracy and get quality data.

[21] Proposed that huge number of sensor devices like RFID is deployed in IoT. The data generated from sensors are mined with the data mining technique like clustering. K-mean is techniques which represent centroid value and define the distance between all object. In, proposed implementation system outlier technique used for increasing the efficiency of data. By using *k*-mean technique we can find out distance with Euclidean distance for centroid data and it can be classified properly. The proposed architecture like link smart middleware architecture uses the upper layer so remaining layer data get filtered. By using McQueen algorithm in K-mean method, which is based on clustering of IoT and big data are used to process the large database in less computation time. Outlier detection used in network for converging traffic in IoT, reduces overhead to analyze raw data and scalability when number of client increases.

3.2. Uncertainty of data

[22] Discussed about one of the challenging issues *i.e*. uncertainty. Author proposed data mining rule called Mining high utility item set to overcome the uncertainty. For example user playing song track depends on the maximum interest and it means, the value of utility is high. There are heterogeneous massive data collected in the form of uncertain data proposed by Chui et al(2007). Generally one item is considered in one transaction, which defines random variable and ignore utility of each item. So this is the main problem in MHUI (Mining High Utility Itemset) over uncertain database. Hence, to overcome this problem, by using UHUI (Uncertain High Utility Itemset) Apriori data mining algorithm, the knowledgeable itemset is obtained. According to experimental result, high probability in itemset gets minimum utility count as 11.05. The downward closure rule kept in UHUI Apriori algorithm for overcoming the uncertainty. Formula are given to reduce the search space TWU(X)= $\sum^{1}_{X \subseteq = Td \wedge Td} \in_{D} Tu(Td)$, hence maximum support count calculated as (44.2x25%) = 11.05. This proposed algorithm provided better performance than direct traversal run time.

3.3. Inaccuracy of data

[23] According to author traditional Apriori algorithm works only in single dimension for valuable information. In array based Apriori algorithm it having many difficulties like calculating itemset in which only one scan is done instead of scanning entire database. So it can improve the performance in speech of search. But it is having defects on unnecessary connection of operation. Array based Apriori algorithm scan only one database, it is considered binary association rule. Each transaction is divided into blocks and stored into stack. Traditionally K item set connect with next item set K + 1 repeatedly upto K – 1. So, this is used to reduce the connection setup but the data received was inaccurate.

After analyzing the customer attribute it ensures the clean data for improving the quality of data and removes the data which is inaccurate. So, for finding the accuracy, from the example like insurance company which is having many itemset and are not duplicate? After dividing into two group {*c*1, *c*2, *c*3}, {*c*4, *c*5, *c*6} find the minimum support degree. In this proposed algorithm array are initialized and shows the total number of "1" in the row. If the number of "1" appears less than 2 times then do delete the frequent set. As the result of array Apriori algorithm, it defines connection process and assigned ascending order to make less number of candidate itemset and increase data. Apriori algorithm used for segmentation and it finds the most relevant item set and deletes non relevant item set.

3.4. Frequently updated time stamped structure

[24] Proposed distributed data mining over centralized data mining technique. Existing technique having 3 layers such as data processing, data gathering and data mining in centralized system. It was very difficult to find frequently used data from huge amount of data. In distributed model, data gathering is done by fog node, data processing by mining technique. Data mining is divided into functional block according to cluch rosser theorem *dma* = *fnofn* – 1 *o fio*….of 1 = *fn*(*d*, *fn* – 1(*d*,….*fi* (*d*, …*f*1(*d*, nil))).

Proposed actor model in distributed computational platform solve computation closer to the data problem. This problem occurs in centralized model. Actor will work at every layer as fog node or data processing layer, so clustering is used for processing. K-mean = find cluster o int cluster, where, find for defining centroid only and int is centriod by random value. By experimental result it describes azure for finding out the execution time more in centralized than distributed. The execution time is faster than centralized because in centralized no need to store but distributed model is used to decrease network traffic on internet. The functional block mapped into actor for analysis, the computing to the network layer of IoT with clustering in smart home, classification for checking the routing path of driver, frequent pattern depend on behaviour between agent and customer. In this system, tools like azure Microsoft, apache mahout are used for implementation at different layers in architecture for IoT.

3.5. Massive real time data

[25] Proposed Traditional database to maintain large frequent itemset into main memory which was not fitting properly. So, by using MapReduce method it divided into level wise breadth first searching algorithm called as Apriori algorithm. While searching predict whether the item set is present for shopping cart using Apriori algorithm, then observed whether it was less efficient in execution time and memory requirement or not. In shopping cart application, it consists two algorithm which is dist éclat and Big FIM for overcoming this problem. This technique does not use recursion for generating frequency itemset, hence it takes less time to execute path between itemset. Dist Eclat works by 3 steps a)finding the frequent itemset b)K-FI generation. *c*) sub tree mining. Dist Eclat increase the processing speed based on time by K-FI with apriori algorithm and mined frequent itemset. Big FIM focus on mining from large database into mapper.

[26] Proposed that Internet is very crucial for every human in society from teenager to parents. Most of the time they spend with internet and for new generation of parents it is very important to take care the behaviours of teenager while using internet because they teenagers lack in making appropriate decision. By using data mining technique, association rule is used between new generation parents and behaviour of teenager. Association rule is applied for parents and k-mean, clustering, DBSCAN rule for teenager behaviour. Making cluster for parents and teenager after applying association rule to find out parental effect on the behaviour of teenager by using Apriori algorithm. Hence, as result of k-mean clustering it shows the best performance to find the use of internet by teenager.

4. OPTIMIZATION IN IOT

Real time data is stored day by day through Internet of Thing with continuous and noisy data. Data management is broad area in intranet as well as in Internet of Things. So it is necessary to reduce communication overhead and storage in architecture of data management. Data is managed between machine to machine, object to object, machine to object, human to machine. Data's are managed like storing, querying, processing in centralized database as offline method. But, in Internet of Thing data is managed by storing, retrieving, querying as online and offline as suggested by [27].

In IoT objects are in heterogeneous and no matter how end points are involved for scalability. For analysis unstructured data classification and indexing method will apply to get flexibility and scalability. Analysis is required to deliver right data in proper place with accurate prediction. After applying analysis method knowledge data is sent to centralized database and then applying MapReduce for less storage space as well as optimized communication overhead proposed by [28].

Nowadays big data used as a fast data processing for large scale industries. So for processing and managing large database, MapReduce gives efficient solution in Internet of Things than the traditional database.

Hadoop is used for implementing MapReduce and distributed file system. The MapReduce is used for distributed file system and run into large cluster environment. It is considered as two phase map and reduce for processing and each phase having key value for input and output. The MapReduce is enabled with Application Programming Interface for using their functionalities as implied by [29].

[30] signified few things, In traditional database it has problem for storing massive volume of data. It is not possible to monitor periodically, so communication overhead and storage data are huge constraint in database for this reason power efficiency also not improved. So, green IoT provide the solution to divide the life cycle into two parts such as front layer and back layer. Communication analysis happened in the front layer and storage data kept in the back layer. Due to monitoring huge database in IoT, power consumption is a major problem. Hence, duty cycle control method is used for optimizing power .

[31] proposed the concept of edge mining and it indicates that the data mining techniques are applied on raw information and relevant information are obtained. In edge mining, its defines state link sense, sleep, transmit event detector and state estimator for optimizing energy. The communication overhead also reduced by using prediction method like STP, Linear STP.

[32] Tipped off about Millions of massive data are collected per week and queries are deployed for analyzing and tracking the data. So efficiency and performance is the challenging issue due to sensor data stored in database. Generated delay response for massive data and its effect on application makes the system to fail. Hence Sea cloud database used to reduce the data overhead and manage the data in cloud. RDB-KV is defined as data type for representing data. But this technique is not helpful to manage huge data for spatial data.

[26] Defined that smart cities compromises smart building planning, mobility, technology for citizen and more. WiFi plays important role for that. Everyday large amount of data used for business purpose and managing it is a main constraint, so big data is used with three steps velocity, volume and variety. It is difficult to store data in fixed amount of time so MapReduce technique used for processing huge data with pair of values by providing different architecture. Also, solved security constraint.

[18] In traditionally monitoring centre produce predictive analysis in the descriptive form for patient and it was too slow for urgent basis treatment like heart attack. In IoT, Mobile health system is used for predicting personal life expectation as well as health status. mHealth is having inference system model to predict accurate information with the help of big data and cloud. All the personal and health information of each patient gathered is a huge data, generated by wearing device like WBAN. Predictive analysis used with the concept of Bayes rule to update the probability of accuracy and it predicts past and future diagnosis, so patient take immediate action on that. By using Bayes rule with Big data we can overcome overhead of large data, frequency of workload is reduced and optimized intelligent data is produced.

[30] alluded that, In previous methods, it was very difficult to manage huge data for IT industry, automotive industry, manufacture industry in the terms of performance, security and large number of manpower. Data stored in database are done by tightly coupled technique. So to overcome this IoT is implemented, in IoT mobile device and sensor node communicates intelligently for increasing digitalization. For maintaining the quality in huge manufacturing and IT industry, YARN is introduced in Hadoop for ecosystem in environment monitoring without user interaction. Day by day huge number of data stored by Hadoop for accurate prediction of data and performance in their function by using different tool in vehicle manufacturing industry, vehicle to vehicle communication, safety warning , traffic prediction etc.

Application for data mining used with IoT is [33]

1. Smart Cities
2. Health Monitoring
3. Environment Monitoring
4. Manufacturing Industries
5. Military surveillance
6. Disaster Monitor

In Fig. 3. indicates that IoT works with data mining techniques and KDD process. In IoT, application worked in real time application through KDD process. In this process described the sensor device sensed data by IoT device then data mining techniques are applied for generating knowledgeable data or pattern from any application like Health Monitoring etc. Today, huge number of real time application based on IoT with data mining techniques for optimized data.

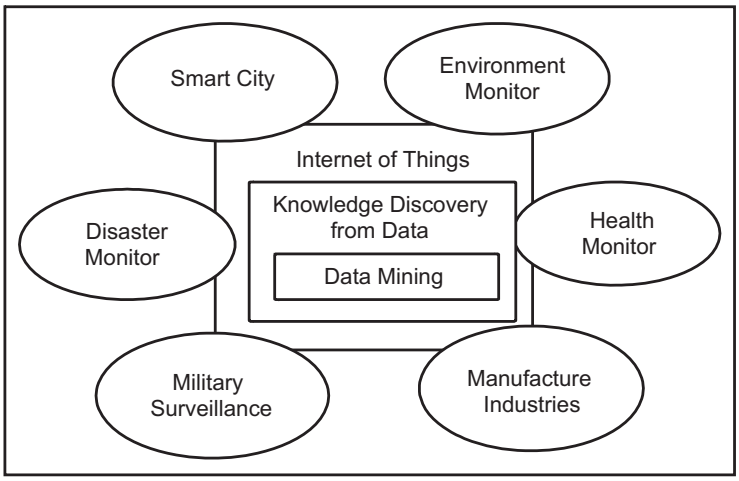

Figure 3: Representation the application of IoT with Data Mining

5. CONCLUSION

Internet of thing is used in every application and it is implemented by different steps of KDD and gets the optimized knowledge by MapReduce technique. Data mining technique is used for many applications like social or personal life. There are many methods like clustering, Apriori algorithm, association rule defined for increasing the performance of communication, scalable, flexibility, low cost and analyze the processing of data. Knowledge discover for optimizing power industry, smart cities, manufacturing industry by different tools like Hadoop.

MapReduce is done by two steps. First is to map the data and then reduce the huge data with pair of value. Some challenging issues are defined like uncertainty, inaccuracy, and massive data in data mining which is overcome by different algorithm technique. Knowledge discovered as pattern value from data mining technique, processing, transforming, cleaning are applied in huge database which are used by IoT sensor devices for storing data.

As we know that big data is used for fast processing in large manufacture industry, IT industry, Environment monitoring, smart cities for real time application. Hence this large data processed by data mining technique and mapper by Hadoop for optimized visible knowledge. In database those data are mined that are not all useful for application so only knowledgeable data visible by MapReduce technique.

REFERENCES

[1] AimadKarkouch, Hassan Al, Mote Assime, Hajar Mousanif, and Thomas Noel,"Data Quality Enhancement in Internet of ThingsEnvironment, IEEE, pp. 1-8,2015.

[2] Chandra Mohan, B., Sandeep, R. and Sridharan, D. "A Data Mining approach for Predicting Reliable Path for Congestion Free Routing using Self-Motivated Neural Network", The Ninth ACIS International Conference on Software Engineering; Artificial Intelligence; Networking and Parallel/Distributed Computing, Thailand (Awarded as Best Paper), Studies in Computational Intelligence (Vol. 149), Springer-verlag, pp. 237-246, 2008.

[3] Chandra Mohan, B. and Baskaran, R. "Reliable Barrier-free Services in Next Generation Networks", International Conference on Advances in Power Electronics and Instrumentation Engineering, Communications in Computer and Information Science (Vol. 148), Springer-Verlag Berlin Heidelberg, pp. 79-82, 2011a.

[4] Chandra Mohan, B. and Baskaran, R. "Energy Aware and Energy Efficient Routing Protocol for Adhoc Network using Restructured Artificial Bee Colony System", International Conference on High Performance Architecture and Grid Computing, Communications in Computer and Information Science (Vol. 169), Springer-Verlag Berlin Heidelberg, pp. 480-491, 2011b.

[5] Chandra Mohan, B. and Baskaran, R. "A Survey: Ant Colony Optimization based recent research in various engineering domains" Expert System with Application, Elsevier, Vol. 39, No. 4, pp. 4618-4627, 2012.

[6] Chandra Mohan, B. "Restructured Ant Colony Optimization routing protocol for next generation network", International Journal of Computer Communication and Control, Vol.10, No.4, pp.493-500, Agora University Press, 2015.

[7] Taeho Hong, and Ingoo Han,"Knowledge based data mining of new information on the internet using cognitive map and neural network",Elsevier,Vol. 23, No. 1, pp.1-8, 2002.

[8] Shuliang Wang," Spatial Data Mining under Smart Earth",IEEE International Conference on Granular Computing, pp.717 – 722, 2011.

[9] Sulayman K. Sowe, Takashi Kimata, Mianxiong Dong, and Koji Zettsu,"Managing Heterogeneous Sensor Data on a Big DataPlatform: IoT Services for Data-intensive Science", IEEE 38th Annual International Computers, Software and Applications Conference Workshops, pp. 295-300, 2014.

[10] Antonio Jara J., Miguel A. Zamora, and Antonio F. Skarmeta,"Knowledge acquisition and management architecture for mobile and personal Health environments based on the Internet of Things", IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications, pp. 1811-1888, 2012.

[11] CarlosGamarra, JosepM.Guerrero,and EduardoMontero ,"A knowledge discovery in databases approach for industrial microgrid planning",Elsevier,pp.615-630, 2016.

[12] Charith Perera, and AthanasiosV. Vasilakos ," A knowledge-based resource discovery for Internet of Things",Elsevier, pp. 1-15, 2016.

[13] Sandeep Singh Rawat, and Lakshmi Rajamani," Probability Apriori based Approach to Mine RareAssociation Rules", 2011 IEEE 3rd Conference on Data Mining and Optimization (DMO), Selangor, Malaysia, pp. 253-258, 2011.

[14] Huan Liu, and Lei Yu ,"Toward Integrating Feature SelectionAlgorithms for Classification and Clustering, IEEE Transaction on Knowledge and data Engineering, Vol. 17, no. 4, pp.491-502,2005.

[15] Chun-Wei Tsai, Chin-Feng Lai, Ming-Chao Chiang, and Laurence T. Yang ,"Data Mining for Internet of Things: A Survey",IEEE Communication Survey & Tutorials, Vol. 16, No. 1, pp. 77-97, 2014.

[16] Yuqing Lan,Yang WangYanni, WangShengwei, and YiDan Yu,"Mining High Utility Itemsets over Uncertain Databases",International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery, pp. 235-238, 2015.

[17] Shen Bin, Liu Yuan, and Wang Xiaoyi ,"Research on Data Mining Models for the Internetof Things",IEEE, pp. 1-6, 2010.

[18] Onder Ondemir A., and Surendra M. Gupta ,"Quality management in product recovery using the Internet of Things: An optimization approach ,Elsevier, pp. 491-504, 2014.

[19] Yongrui Hua Wangc, Athanasios V. Qina, Quan Z. Shenga, Nickolas J.G. Falknera, and Schahram Dustdarb Vasilakos,"When Things Matter: A Survey on Data-Centric Internet of Things", Journal of Network and Computer Applications , pp. 1-20, 2016.

[20] Alberto M. C. Souzaa, Jos´e R, and Amazonasb A ,"An Outlier Detect Algorithm using Big Data Processing and Internetof Things Architecture", International Workshop on Big Data and Data Mining Challenges on IoT and Pervasive Systems, pp. 1010-1015, 2015.

[21] Chui K.C., Kao B., and Hung E., "Mining frequent itemsets from uncertain data," in Springer , pp. 47–58, 2007.

[22] Liu, Jun Liu , and Meiling Cong ,"The Optimization of Apriori Algorithm Based on Array And Its Application in The Analysis of Insurance Clients",IEEE conference, pp. 58-61 2015.

[23] Ivan Kholod, Mikhail Kuprianov, and Ilya Petukhov ,"Distributed Data Mining Based on Actors for Internet of Things", IEEE conference, pp. 480-484, 2016.

[24] Kiran Chavan, Priyanka Kulkarni, Pooja Ghodekar, and S.N. Patil,"Frequent Itemset Mining for Big Data",IEEE, pp. 1365-1368, 2015.

[25] Chonnikarn Rodmorn, Mathuros Panmuang, and Khuanwara Potiwara,"Analysis of the Internet Using Behavior of Adolescents by Using Data Mining Technique", 7th International conference on Information Technology & Electrical Engineering,Thailad, pp.398-402, 2015.

[26] Anup W. Burangel, and Harshal D. Misalkar ,"Review of Internet of Things in Development of Smart Cities with Data Management &Privacy ", 2015 International Conference on Advances in Computer Engineering and Application (ICACEA)IMS Engineering College Ghaziabad, India, pp.189-195, 2015.

[27] Najah Abu Ali , and Mervat Abu-Elkheir," Data Management for The Internet of Things: Green Directions" ,GC'12 Workshop: Green Internet of Things, IEEE, pp.386-389, 2012.

[28] Elena I. Gaura, James Brusey, Michael Allen, Ross Wilkins, Dan Goldsmith, and Ramona Rednic," Edge mining the Internet of Things",IEEE, pp. 1-10, 2012.

[29] Zhiming Ding, Qi Yang, Hong Wu," Massive Heterogeneous Sensor Data Management in the Internet of Things", 2011 IEEE International Conferences on Internet of Things, and Cyber, Physical and Social Computing, pp. 100-108, 2011.

[30] Andre Luckow, Ken Kennedy,Fabian Manhardt, Emil Djerekarov, Bennie Vorster,and Amy Apon,"Automotive Big Data: Applications, Workloads and Infrastructures", IEEE International Conference on Big Data (Big Data), pp. 1201-1210, 2015.

[31] Welbourne E., Battle L., Cole G., Gould K., Rector K., and Raymer K., "Building the Internet of Things Using RFID: The RFID Ecosystem Experience Internet Computing", IEEE, Vol.13, No. 3, pp. 48-55, 2009.

[32] Daniele Miorandi, Sabrina Sicari, Francesco De Pellegrini, and ImrichChlamtac, "Internet of things: Vision, applications and research challenges",Elsevier, pp. 1497–1516, 2012.

[33] YuxiLiu , and Guohui Zhou ,"Key Technologies and Applications of Internet of Things", Fifth International Conference on Intelligent Computation Technology and Automation, pp.197-200, 2012