ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

Identifying the Useful Data in the Repository to Reduce Data Redundancy and Effective Utilization of Data, Using IOT

Uday Shankar Sekhar, Gopinadh Sasubilli

EY, Senior Consultant

Mc Donald's, IL, Sr. Solution Architect

ABSTRACT: Data is being created from different sources over the globe and most of the data is in unstructured format and we are in confusion that which data is to keep and which has to exclude from the repository. Because of maintenance of all the format data we are creating big data and its creating a huge issue for data maintenance and also to provide security to the repositories like data centers and also cloud repositories. Here we are providing the method to identify the kind of data we can keep and maintain and what the data we need to discard from the repository. When we consider weather prediction system we need to have past data and that will be in different formats but we don't need complete data from the repository and we can discard some part of data permanently which will be helpful for the researchers to maintain the data with an efficient manner. This scenario includes mining the repository based on our requirement and implementing some machine learning techniques to identify which data is most valuable and needed to perform some of the research operations like designing the prediction model and identifying a novel approach in a specific domain. This research can be implemented in any domain which is having more amounts of big data. Health care domain is the platform which was used more to perform any mining activities and to identify a novel thing in the real world. Big Data issue has to be resolved in medical domain which leads to the good architecture of maintaining data and also securing data is also possible in an efficient manner.

KEYWORDS: Mining, Heath care, Machine Learning, Big Data, IoT, repository, Cloud repository.

I. INTRODUCTION

Mining the data for designing and useful model is a good sign of research work and the main issue here is to maintain the data which is not useful and also it leads to security bleach. Here we are proposing a new approach which will identify the data weightage and which data is needed to be included in the repository and which we need to discarded. Here we are considering a real time scenario as example, maintenance of IoT. Consider any activity based on the social networks [1] which deals with the understanding the behavior of a group of peple in social network and it is very difficult to gather the data relevant to the specific domain regarding the social networking [2]. For example if we need to consider any social networking consider twitter for mining the tweets of the group of people who are discussing on a specific topic. Based on the tweets of that trending topic we can perform some prediction model but here the important thing is if we consider only the trending tweets we cannot perform effective mining operation, because on the same concept any other outer member from other domain also can tweet and we may miss this kind of relevant data.

The Internet of things (IoT) is the system of physical gadgets, vehicles, home apparatuses, and different things installed with hardware, programming, sensors, actuators, and system availability which empowers these articles to interface and trade data.[1][2][3] Each thing is interestingly identifiable through its implanted figuring framework yet can between work inside the current Internet foundation.

Specialists gauge that the IoT will comprise of around 30 billion questions by 2020.[4] It is likewise evaluated that the worldwide market estimation of IoT will reach $7.1 trillion by 2020.[5]

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

The IoT enables items to be detected or controlled remotely crosswise over existing system infrastructure,[6] making open doors for more straightforward combination of the physical world into PC based frameworks and bringing about enhanced effectiveness, exactness and financial advantage notwithstanding diminished human intervention.[7][8][9][10] When IoT is expanded with sensors and actuators, the innovation turns into an occasion of the more broad class of digital physical frameworks, which likewise includes advancements, for example, brilliant networks, virtual power plants, savvy homes, insightful transportation and keen urban communities.

"Things", in the IoT sense, can allude to a wide assortment of gadgets, for example, heart checking inserts, biochip transponders on cultivate creatures, cameras gushing live sustains of wild creatures in waterfront waters,[11] cars with worked in sensors, DNA examination gadgets for
natural/sustenance/pathogen monitoring,[12] or field operation gadgets that help firefighters in hunt and protect operations.[13] Legal researchers propose in regards to "things" as an "inseparable blend of equipment, programming, information, and service".[14]
These gadgets gather valuable information with the assistance of different existing advances and afterward self-governing stream the information between other devices.[15]

Starting at 2016, the vision of the Internet of things has advanced because of a meeting of different innovations, including pervasive remote correspondence, continuous investigation, machine learning, item sensors, and installed systems.[13] This implies the conventional fields of inserted frameworks, remote sensor systems, control frameworks, mechanization (counting home and building computerization), and others all add to empowering the Internet of things.[15]

The idea of a system of shrewd gadgets was talked about as ahead of schedule as 1982, with an altered Coke machine at Carnegie Mellon University turning into the primary Internet-associated appliance, ready to report its stock and whether recently stacked beverages were cold. Mark Weiser's original 1991 paper on pervasive processing, "The Computer of the 21st Century", and in addition scholarly scenes, for example, UbiComp and PerCom created the contemporary vision of IoT. In 1994 Reza Raji depicted the idea in IEEE Spectrum as "[moving] little parcels of information to a huge arrangement of hubs, in order to incorporate and robotize everything from home apparatuses to whole factories". Between 1993 and 1996 a few organizations proposed arrangements like Microsoft's at Work or Novell's NEST. In any case, just in 1999 did the field begin gathering energy. Bill Joy imagined Device to Device (D2D) correspondence as a major aspect of his "Six Webs" system, introduced at the World Economic Forum at Davos in 1999.

The idea of the Internet of things wound up noticeably mainstream in 1999, through the Auto-ID Center at MIT and related market-examination publications. Radio- recurrence recognizable proof (RFID) was seen by Kevin Ashton (one of the originators of the first Auto-ID Center) as an essential for the Internet of things at that point. Ashton lean towards the expression "Web for things." If all items and individuals in day by day life were outfitted with identifiers, PCs could oversee and stock them. Besides utilizing RFID, the labeling of things might be accomplished through such innovations as close field correspondence, scanner tags, QR codes and advanced watermarking.

This research work is completely based on understanding the importance of the available data and how to evaluate the required data and how to discard which is not relevant. Here we need to identify the domain before performing some operations. Let’s make neural networks are the mining tool for our domain of research[3-5]. Using ANN we need to plot some graphs in the heterogeneous space and categorize the variables based on their impact factor in the model. Based on the high impact factor in the model we need to implement multiple linear regressions forward elimination process. Once we perform this operation in machine learning we can identify the variables which are highly having impact on the current model and we can discard remaining things from the repository permanently. In this regard we need are considering a specific domain of implementation for better understanding of the concept and further sections in this articles describes the remaining things related to this research which will decrease the load on the repository and avoid the traffic in the network for data transmission. When we trying to decrease the big data which is not useful we can release the cost investment on the irrelevant data so that we can perform better security measures to the required data only.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

Examiners concur that IoT is relied upon to become enormously finished the following decade and it will affect each industry. We trust this galactic development is driven because of quick positive business affect conveyed from another age of reasonable sensors; versatile and secure cloud framework; and progressed investigation. The worldwide addressable market for IoT is assessed to top $1.3 trillion by 2019, making new business openings and making associations more quick witted about their items to enable them to give better client benefit. We consider this open door the Quantified Economy.

As indicated by Gartner, "43% of associations are utilizing or plan to actualize IoT this year," which means organizations comprehend the advantages and significance of IoT, and are making a positive advance forward to receive the innovation. There are a couple of early industry adopters like armada administration associations, oil, utilities, gas, and assembling that are upsetting their operational foundation to concentrate on cost investment funds, upgraded use of benefits and general efficiencies.

Because of potential picks up in vitality productivity and preservation, government and regions in real urban areas are likewise exploiting associated gadgets by attaching savvy meters to building administration frameworks. You can interface your gadget through reasonable sensors, that give imperative information from many sources. For instance, a building director can track constant vitality or water use by envisioning the information in a dashboard. By just observing the utilization progressively and looking at it month over month, utility expenses can go down 5-7%. Past that, these same frameworks give over the air and robotized control of building administration framework parts.

In any case, before you concentrate on IoT information accumulation, there are a couple of ventures to take to comprehend the esteem you can get from the information.

Further sections will consists of domain specific discussion, later how to implement the concept, what are the issues in this process, estimated result and conclude the discussion with sample future scopes and further research implementations.

II. DOMAIN SPECIFIC APPROACH

As we discussed in every sector of research we have a scope of reducing the big data issue. For suppose in this article we would like to discuss about the health care domain and privacy preserving of the patients who are using IoT device as EHR data[6-7]. We have a patient with IoT device connected with him and his personal doctor and it will work based on that patient activities. Whenever he used that band it will instruct the patient to do some operations like run, walk take water, take limit food etc. But when the patient was dead and that information is no where require to be maintained. Here we are creating a huge amount of big data with this kind of operations using these IoT devices. This is the case of only one person. If the case is continued with a group of people and what if all are dead and their data is maintained in the repository uselessly. Here we may get an issue by removing that kind of data. Researchers may think that removed data may be useful for their research. But we can judge whether that data can be useful for further process or not by using forward elimination process of data mining. Using this process we need to identify the variables which are needed to be in the model. But doing this kind of operations we can eliminate the big data effect on the storage and the processing of the request and we can effectively maintain the things in the data processing.

We are proposing a domain specific architecture to identify the data which is not required and can be eliminated. This architecture will assume that a device which is working based on the human heart rate. That patient can store his/her personal, medical , bank etc information into that device. Lets discuss the architecture of this IoT device related to big data in detail in next section.

III. IMPLEMENTATION AND ITS PROPOSED ARCHITECTURE RESULTS

Consider a device with connecting sensor in it which will be continuously connected with human heart. In that device we are storing all the medical related information, personal information, picture, messages, reports, bank operations etc.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

When the patient starts wearing this he will be continuously monitored by that device regarding his health condition and will regularly update in the server time to time. Suppose we consider a pace maker to control heart patient, it will record that patients information time to time and update it to the server. What happens to that data when that patient expires? The data will be alive in the server for long years which is a very big burden to maintain. Likewise we need not to maintain that kind of huge data which is irrelevant and can be terminated with its services. In this architecture we are designing a simple architecture by considering the prescribed IoT device [8-10] is waterproof and need not to be removed at any cost.

As this is considered as device works with heart beat it will continuously records the heart functionality but its not to check the heart capacity but It checks to keep working. If the heart beat stops then the device will automatically terminates its working. But the data it recorded it now will be stored in the repository as well as in the device cache memory.

That data is supposed to be deleted when the patient is dead as this is not required. An IoT can only use oximeter to read oxygen levels of the patients, cannot check BP , Sugar levels of the patients by fixing it to external body. So there will be no big information available with it so we can terminate maintenance of the patience device and also to the data.

In this architecture we will maintain two kinds of repositories which the device will be connecting and when the patient start using it, it will start recording the heart functionality using sensors and whenever the heart functionality stops immediately sensor will report to the server that heart beat stops and we can terminate the service of that patient. There are two ways to decide whether the heart beat was stopped or not. 1) When the patient's heart beat is not recorded for maximum of 5 seconds then the algorithm on which the device is working will terminate automatically. 2) When heart rate fallen down after reaching a maximum point and stood at zero that means the person is no more. Our algorithm will decide based on the two criteria's and based on that it will terminate the services. Whenever services are terminated automatically algorithm will delete all the patient related data from the server so that it will continuously helps to avoid big data[11-15].

By removing this kind of data we can provide the security to the patient's data and avoids security bleach to the data.

Here we can see all the personal information of the person. Even his travelling data, like tourism etc, everything is connected to server through a mobile application which will be further connected to the server which is remotely hosted by the admin.

The main motto of this kind of architecture is to identify and eliminate unwanted data from the server and to reduce the server burden and reduce the storage capacity. Using this kind of architectures we can avoid the security bleach.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

Data	Type	Characteristics	Examples
Vitals	Numerical, Temporal	Typically measured every second/minute within ICU and every few hours outside ICU	Blood Pressure, Respiration Rate, Heart Rate
Lab Tests	Numerical	Typically measured a few times, investigation depends on patient's condition and diagnoses	Blood Glucose, Uric Acid
Medication Orders	Numerical, Temporal	Physician orders of prescribed medications	Insulin, Aspirin
Procedures	Numerical, Temporal	Medical/surgical procedures performed on the patient	Craniotomy, endoscopy
Diagnoses	Numerical, Temporal	Diagnoses of past and current conditions	Sepsis, Diabetes
Nursing Notes	Text, Temporal	Assessment of patient's condition including subjective observations	See Figure 3
Radiology	Image, Text	Radiology images accompanied by reports from the radiologists	X-Ray, CT Scan
Demographic	Numerical, Static	Demographic details of the patients	Age, gender, ethnicity

Figure 1: Expected Data set which is reduced in the EHR model.

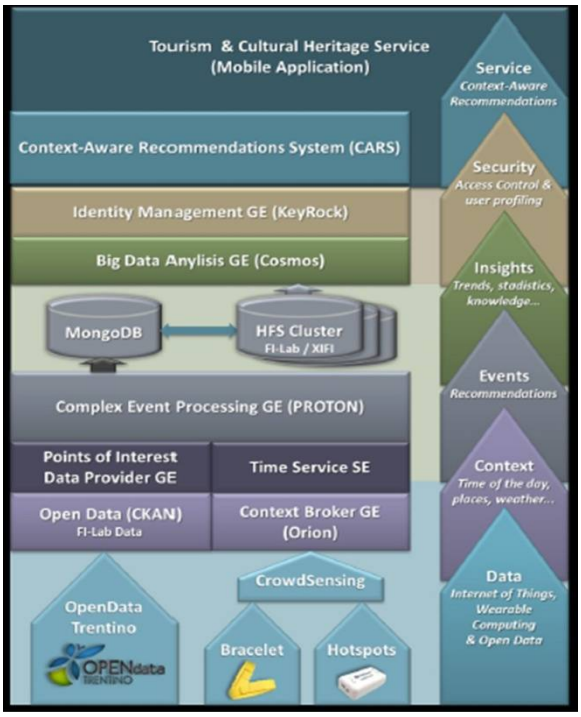

Figure 2: Sample architecture of IoT device connected with server.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

IV. ADVANTAGES

There are few advantages list for this kind of approach and they are as follows:

i. Effectively gather patients information with less funding.
ii. Cost effective EHR maintenance is possible.
iii. Better understanding of the IoT architecture and we can transfer data in an efficient manner
iv. High level sensors are utilized to get apt information from the device.
v. Can perform some prediction models if we want.
vi. Using this data we can perform forward elimination to get information regarding the valid variables which are needed to be included into the model and also to decide which variables data can be stored.
vii. Better utilization of data mining techniques is possible
viii. Architecture can be refined if we need, because it's a generic architecture.
ix. Cost effective
x. Easily shared with the consult doctors if they needed
xi. Easy to carry the medical records instead of missing anything or misplacing any one record.
xii. Can be stored to local storage if needed.
xiii. All types of formats of data can be generated based on the requirement.

These are few advantages which are considered to be discussed.

V. DISADVANTAGES

As we have advantages we also have downsides of every architecture. Let's discuss them in brief.

i. If once data is removed it cant be revert back
ii. Sometimes we may face poor performance because of the network issue.
iii. Once network was lost while transferring the data we may or may not revert back that data and we cannot trace out how much data was transferred and how much is pending,
iv. As we are considering all type of formats of data sometimes it may lead to difficulty in converting the data into required format. For example we may find issue with transforming image into text format to get hidden information in that image.

VI. FUTURE SCOPE OF RESEARCH

Considering the present research on the EHR data and utilizing IoT devices are means of communications and data transfer we can perform better mining techniques to identify the data which is more required to perform the prediction model and which can be eliminated from the server as the patient. Dead. If the case is like this " If the data is removed from device immediately after the patient death from the device, the same data must be available in the server and the mobile application it is connected with because we cannot erase information before validating it. In future work of this kind of research in IoT for better utilizing of the resources and reducing big data requires some better architecture which will validate the data in recurrent manner which refines the data and decide whether it can be stored or eliminated.

VII. CONCLUSION

In this article we discuss about the method we can gather information regarding the patient, not only medical also the personal and identifies how we can share, transfer and communicate data among the devices and how to manipulate data according to our requirement. We have seen the process of managing data to eliminate the issues with big data and we can efficiently utilize this kind of architecture for better utilization of EHR data and use for further research.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 7, Issue 5, May 2019

REFERENCES

[1] XI XIONG *, “Empirical Analysis and modeling of the activity dilemmas in Big Social Networks”, Oct 2016.
[2] F. C. Santos and J. M. Pacheco, ``Scale-free networks provide a unifying framework for the emergence of cooperation," Phys. Rev. Lett., vol. 95, p. 098104, Aug. 2005.
[3] K. B. To and L. M. Napolitano, ``Common complications in the critically ill patient," Surgical Clinics North Amer., vol. 92, no. 6, pp. 1519_1557, 2012.
[4] C. M. Wollschlager and A. R. Conrad, ``Common complications in critically ill patients," Disease-a-Month, vol. 34, no. 5, pp. 225_293,1988.
[5] S. V. Desai, T. J. Law, and D. M. Needham, ``Long-term complications of critical care," Critical Care Med., vol. 39, no. 2, pp. 371_379, 2011.
[6] N. A. Halpern, S. M. Pastores, J. M. Oropello, and V. Kvetan, ``Critical care medicine in the United States: Addressing the intensivist shortage and image of the specialty," Critical Care Med., vol. 41, no. 12, pp. 2754_2761, 2013.
[7] A. E. W. Johnson, M. M. Ghassemi, S. Nemati, K. E. Niehaus, D. A. Clifton, and G. D. Clifford, ``Machine learning and decision support in critical care," Proc. IEEE, vol. 104, no. 2, pp. 444_466, Feb. 2016.
[8] O. Badawi et al., ``Making big data useful for health care: A summary of the inaugural MIT critical data conference," JMIR Med. Informat., vol. 2, no. 2, p. e22, 2014.
[9] C. K. Reddy and C. C. Aggarwal, Healthcare Data Analytics, vol. 36. Boca
Raton, FL, USA: CRC Press, 2015.
[10] D. Gotz, H. Stavropoulos, J. Sun, and F. Wang, ``ICDA: A platform for intelligent care delivery analytics," in Proc. AMIA Annu. Symp., 2012, pp. 264_273.
[11] A. Perer and J. Sun, ``Matrix_ow: Temporal network visual analytics to track symptom evolution during disease progression," in Proc. AMIA Annu. Symp., 2012, pp. 716_725.
[12] Y. Mao,W. Chen, Y. Chen, C. Lu, M. Kollef, and T. Bailey, ``An integrated
data mining approach to real-time clinical monitoring and deterioration warning," in Proc. 18th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining. 2012, pp. 1140_1148.
[13] J. Wiens, E. Horvitz, and J. V. Guttag, ``Patient risk strati_cation for hospital-associated C. Diff as a time-series classi_cation task," in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 467_475.
[14]"R. T. Mylavarapu and B. K. Mylavarapu, ""Huge information extraction techniques of Data Security,"" 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 179-183.
doi: 10.1109/ICICCT.2018.8473017"
[15] "Kumar. Attangudi P. Perichappan, S. Sasubilli and A. Z. Khurshudyan, ""Approximate analytical solution to non-linear Young-Laplace equation with an infinite boundary condition,"" 2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), Sukkur, 2018, pp. 1-5. doi: 10.1109/ICOMET.2018.8346349"
[16] Attangudi Perichiappan Perichappan, Kumar. (2018). Greedy Algorithm Based Deep Learning Strategy for User Behavior Prediction and Decision Making Support. Journal of Computer and Communications. 06. 45-53. 10.4236/jcc.2018.66004.
[17]"S. Chandrasekaran, ""Contemplated Method for Predicting Disease by Deep Learning Approach Over Big Data,"" 2018 International Conference on Research in Intelligent and Computing in Engineering (RICE), San Salvador, 2018,pp1-5.doi:10.1109/RICE.2018.8509090"
[14] S. Saria, D. Koller, and A. Penn, ``Learning individual and population level
traits from clinical temporal data," in Neural Inf. Process. Syst. (NIPS), Predictive Models Personalized Med. Workshop, 2010.
[15] I. Batal, H. Valizadegan, G. F. Cooper, and M. Hauskrecht, ``A pattern mining approach for classifying multivariate temporal data," in Proc. IEEE Int. Conf. Bioinformatics Biomed. (BIBM), 2011, pp. 358_365.