International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 9, Issue 3, March 2020, ISSN: 2278 – 1323

Survey on Data Management in IoT using Machine Learning Algorithms

T.Kalai Selvi, R.Karunamoorthi, R.Narendran, Dr.G.Saravanan

Abstract - **The Internet of Things (IoT) is a network of items or devices that are connected to the Internet, usually via sensors, and can relate to each other and the data they generate. These connected "things" ranging from smart phones and cars to refrigerators, thermostats, and mirror are gradually incoming every part of our lives. With 41.6 billion connected devices expected by 2025, IoT's stability is only going to increase. In this paper, we have presented the survey about several papers on the Data Management in IoT using Machine Learning Algorithms.**

Index Terms- **Data Management, IOT Applications, Internet of Things (IoT), Machine Learning Algorithms,**

I. INTRODUCTION

IoT implementation has increased significantly in the last five years due to the accessibility of massive computing power, innovations in data-processing technology, and the initiation of machine learning and natural-language processing algorithms. IoT has opened an entirely new arena for customers to address their long-standing issue of connecting devices and using the resulting data to positively control decision-making process. IoT also opens an entirely new variety of use cases where customers can operationalize actions on the IoT devices in real-time something that was not possible a few years ago.

IoT data management helps organization be aware of how environmental conditions and user activities can affect the performance of their products. IoT sensors can also be used to evaluate manufactured goods performance metrics. The data collected by these sensors can be used to improve future versions of products.

***Manuscript received March, 2020*.**

***T.Kalai Selvi**, Assistant Professor (SLG-I), Dept. of CSE, Erode Sengunthar Engineering College, Erode, Tamil Nadu, 9842899992, tkalaiselvi1281@gmail.com.*

***R.Karunamoorthi**, Assistant Professor, Dept. of CSE, Erode Sengunthar Engineering College, Erode, Tamil Nadu, 9025044806, karunamoorthir@gmail.com.*

***R.Narendran,** Assistant Professor, Dept. of CSE, Erode Sengunthar Engineering College, Erode, Tamil Nadu, 9944519941, narenofficial124@gmail.com.*

***Dr.G.Saravanan,** Assistant Professor, Dept. of CSE, Erode Sengunthar Engineering College, Erode, Tamil Nadu, 9894379369, gsaravanan.esec @gmail.com.*

II. IMPLICATION - IOT

Significance of IoT Data Management

The following are the importance of Data Management in IoT:

- How to get back the data from the IoT systems and make it accessible for the analytics systems and for conclusion.
- The capability to consume the data from IoT systems into the data lake. In most scenarios, organization also wants to improve and cleanse the data and the analysts have enriched data for their analytics.

Prospective for IoT data management

Managing data from IoT devices is an important feature of a real-time analytics. The following key ability are needed to handle IoT data demands:

1. Multitalented connectivity and facility to handle data range

IoT systems have a variety of standards and IoT data hold on to a wide range of protocols (MQTT, OPC, AMQP, and so on). Also, most IoT data exists in semi-structured. Therefore, data management system must be able to connect to all of those systems and stay to the various protocols. It is essential that the solution to carry both structured and unstructured data.

2. Edge processing and improvements

A good data management resolution will be able to riddle out flawed records coming from the IoT systems. It should also be able to improve the data with metadata.

3. Big data processing and machine learning

IoT data becomes very large volumes which require the ability to run enrichments so that the data is ready to be consumed in real time. Also, many customers want to put into use ML models so that they can take preventive steps.

4. Concentrate on data flow

Data coming from IoT systems can change over time due to actions such as firmware upgrading. It is important that, data management solution can automatically concentrate on data flow without interrupt the data management procedure.

5. Real-time observing and vigilant

IoT data absorbing and dispensation never stops. Data management solution should provide real-time monitoring with flow visualizations with respect to performance and throughput. The data management solution should also provide alerts in case any issues arise during the process.

III. IOT APPLICATIONS AND ITS ML-ALGORITHMS

Machine learning and IoT:

Machine learning attempt to minimize human being involvement in tasks that can be automated and it is fully related to IoT. Machine learning opens many prospects to computerize and optimize the world of IoT. Using machine learning algorithms, organizations can use IoT data to discover patterns and build models which can then be scored in real time on the IoT data to operationalize the models.

IoT applications with ML algorithms are:

- Smart traffic prediction using classification, anomaly detection, and clustering techniques.
- Energy usage prediction using linear regression, classification, and regression trees.
- Food safety prediction using Naive Bayes algorithm.
- Smart city and smart citizen initiatives with K-means clustering algorithm
- Health care- anomaly detection in recorded medical data
- Insurance – anomaly detection
- Utilities (gas, water) – demand supply prediction

The following table shows the relationship between the IoT application and Machine Learning Algorithms:

IoT Applications	Service provided	Machine Learning Algorithms
Manufacturing	• Remote monitoring and diagnostics • Production line automation • Equipment handling	Anomaly detection in equipment usage and function
Utilities (energy, water, gas)	• Real time collection of usage data • Demand supply prediction • Load balancing	Usage prediction, demand supply prediction
Healthcare	• Remote expert doctor monitoring • Chronic disease management	Anomaly detection
Smart city	• Collect useful data like traffic congestion, energy use, and CO2 levels.	K-means clustering algorithm
Smart traffic prediction	• Regulate the flow of traffic	Classification, anomaly detection, and clustering techniques
Insurance	• Collection of user data • Remote inspection and assessment of damage and accidents	Anomaly detection
Food safety prediction	• Quality detection • Production standards	Naive Bayes algorithm
Consumer goods and retail	• Monitoring of supply chain inventory	Supply chain analytics
Transportation	• Real time vehicle tracking and optimization for logistics and public transportation systems	Prediction, Optimization and decision support systems

IV. LITERATURE SURVEY

[1] In this paper, the following eight algorithms have been analyzed Support Vector Machine (SVM), K-

Nearest Neighbours (KNN), Linear Discriminant Analysis (LDA), NaiveBayes (NB), C4.5, C5.0, Artificial Neural Networks (ANNs), and Deep Learning ANNs (DLANNs). Out of these eight algorithms C4.5, C5.0, ANN and DLANN algorithms achieved far better than SVM, KNN, NB and LDA. The classification accuracy (CA) is very close to each other for C4.5, C5.0 and ANN algorithms. The average accuracy (AAC) for C4.5 is 97.15%, which carry out somewhat improved than 96.61% AAC of the C5.0 algorithm. AAC of ANN is 96.19%. C4.5 tops among all the eights algorithms in terms of the classification accuracy, followed closely by C5.0. SVM shows its limitation towards multi-label data classification as compared to binary classification problem where its performance is one of the best. SVM executed better than KNN with 4.09% higher AAC. The choice of k-neighbours and distance measure aﬀ ects the CA of KNN. The remaining two algorithms NB and LDA, performed the worst in terms of CA.

In terms of execution time, both NB and LDA algorithms are the best among all the eight algorithms. LDA is good than NB. Average processing time (APT) of C4.5, C5.0 is 7.70 and 7.21 seconds respectively. SVM uses a lot of system resources and has slow processing speed. KNN is lighter and has low execution times. ANNs and DLANN have higher computational requirements. For IoT, there can be problems where high CA does not be concerned much, but processing time is considered. In those cases, NB and LDA can be useful.

Based on our preface consideration, DLANNs can have the best CA among all the simulated algorithms. It is experimental that improved classification accuracy could be attained by increasing the epochs, hidden layers and neurons. In DLANNs, CA also depends significantly on its parameters tuning. DLANNs have a very complex structure, need a large amount of system resources, and as a result, DLANN algorithm has the highest execution time among all the eight algorithms. C4.5, C5.0, ANNs and DLANNS can give relatively higher accuracy results.

[2] In this paper, different data mining algorithms like decision tree, k- means clustering and naive bayes are applied to the datasets with the help of open source data science platform Rapid Miner. The datasets are retrieved from "UCI Machine Learning Repository". Hungarian Database is taken as case study under Heart Disease dataset with 14 attributes used that are defining the datasets. Decision tree when run with datasets showed the better result as compared to others. K-means clustering model provides different clusters by dividing datasets into two different clusters based on their attributes while naive bayes algorithm provides classification by classifying them into different classes.

[3] In this paper, a data mining model has been proposed which consists of six layers. The entire data mining model works with an added perception of security which tries to secure the data and keep the privacy of the user intact. This model also deals with the networking of the data and check data is coming from which environment. The data in Internet of Things has numerous features, such as distributed storage, a lot of relevant data on time and place, and limited node resources etc., which make the task of data accessing in IoT challenging. Four data mining models have been examined: multilayered data mining model, distributed data mining model, grid-based data mining model and big data mining model.

[4] In this paper, the system minimizes the overhead caused by repeated data in IoT. A repetitive data management algorithm is used to handle repeated data in IoT. RDMA algorithm is applied on the data collected from the different sensors like DHT11, Digital light sensor, pressure sensor and rain detecting sensor. In this work, comparison between the repeated data and non-repeated data is presented successfully. The proposed algorithm reduces the 44.83% network load and eliminates the data processing overhead due to repeated data generated by different sensors in IoT. The use of MongoDB database results in high availability of data, faster data processing. The proposed algorithm can be adopted easily by the other IoT applications such as Home Automation, Smart Agriculture, Healthcare applications and other applications, since it reduces the overhead incurred data processing.

[5] This paper presents a comprehensive survey on the IoT in large-scale petrochemical plants as well as recent activities in communication standards for the IoT in industries. It also addresses the key enabling middleware approaches, e.g., an industrial intelligent sensing ecosystem (IISE), which allows rapid deployment and integration of heterogeneous wireless sensor networks with advancements in crowdsensing based services. In addition, it highlights the research issues of data management in the IoT for large-scale petrochemical plants. As a part of an industrial intelligent sensing ecosystem that allows participating in crowdsensing, a middleware system integrated with the IoT will emerge to manage several issues associated with data management in large-scale petrochemical plants. Along with current middleware research in large-scale industry, it has been observed that petrochemical industries integrated with the IoT will drive the new growth paradigm. Standardization, real-time update, geo-awareness, and open access are the main issues to be addressed in the future.

[6] This paper presents a data capture system and data management catalog with solutions addressing the challenges of curating IoT data applied to purpose-built

machine learning deployments. The system implements a data curation pipeline where the raw data collected is processed in stages from online collection and insitu labeling, through offline reviewing and (re-)labeling, to data curation for ML model training. This process has been performed in iteration, as the end-user (the ML modeler) can trigger processes in earlier stage depending on his/her needs and outcome from using the curated data.

[7] In this paper, architecture for IOT system have been proposed, that will have the capability of triggering quick reactions based on cell critical data evolutions. The response time is improved by adding: a new layer responsible for Critical Data storing and processing. This layer will use only database primitive data types (numbers) in order to avoid future data processing in the following thread; a new Thread for real time analysis and reaction to critical data evolution. This Thread will only used primitive data types stored in the database; a database system that uses query caching.

REFERENCES

[1] Furqan Alam, Rashid Mehmood, Iyad Katib, Aiiad Albeshri, “Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)” , Procedia Computer Science 98 (2016) 437 – 442, International Workshop on Data Mining in IoT Systems (DaMIS 2016).

[2] Manish Joshi , Bramah Hazela, Vineet Singh, “An application of IoT on Hungarian database using Data mining Techniques: A collaborative approach”, 978-15090-6403-8, 2017 IEEE.

[3] Aashi Singh , Shilpi Sharma , “Analysis on Data Mining Models for Internet of Things” International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud- I-SMAC 2017).

[4] “Repeated Data Management Framework for IoT: A Case Study on Weather Monitoring and Forecasting”, Narasimha Swamy S , Sowmyarani C N, 978-1-5386-3039-6, 2018 IEEE, 4th Int’l Conf. on Recent Advances in Information Technology | RAIT-2018

[5] Lei Shu, Senior Member, IEEE, Mithun Mukherjee, Member, IEEE, Michael Pecht, Fellow, IEEE, Noel, Crespi, and Son N.Han, “Challenges and Research Issues of Data Management in IoT for Large-Scale Petrochemical Plants”, 1937-9234, 2017 IEEE

[6] Keith Grueneberg, Bongjun Ko, David Wood, et.al “IOT Data Management System for Rapid Development of Machine Learning Models”, 978-1-7281-2711-8, 2019 IEEE DOI 10.1109/ICCC.2019.00021, 2019 IEEE International Conference on Cognitive Computing (ICCC).

[7] Catalin Constantin, Claudia Monica, “Large Data Management in IOT applications”, 978-1-4673-8606-7-2016 IEEE.

T.Kalai Selvi received M.E degree in Computer Science and Engineering from Mahendra Engineering College, Namakkal and presently she is working as Assistant Professor (Selection Grade-I) in Department of Computer science and Engineering at Erode Sengunthar Engineering College, Erode. She has published 6 papers in reputed Journals and more than 10 papers in various National and International conferences. Her Research interests include Cloud Computing, IOT. She is a Life member of ISTE and CSI.

R.Karunammorthi received M.E degree in Computer Science and Engineering from Jayam College of Engineering and Technology, Dharmapuri and presently he is working as Assistant Professor in Department of Computer science and Engineering at Erode Sengunthar Engineering College, Erode. He has published 6 papers in reputed Journals and more than 10 papers in various National and International conferences. His Research interests include Big Data, IOT. He is a Life member of ISTE and CSI.

R.Narendran received M.E degree in Computer Science and Engineering from Nandha College of Technology, Erode and presently he is working as Assistant Professor in Department of Computer science and Engineering at Erode Sengunthar Engineering College, Erode. He has published 2 papers in reputed Journals and more than 6 papers in various National and International conferences. His Research interests includes IOT. He is a Life member of ISTE.

Dr.G.Saravanan received M.E degree in Computer Science and Engineering from Government College of Technology, Coimbatore and presently he is working as Assistant Professor in Department of Computer science and Engineering at Erode Sengunthar Engineering College, Erode. He has published more than 10 papers in reputed Journals and 10 papers in various National and International conferences. His Research interests includes Networking. He is a Life member of ISTE, IAENG and IFERP.