DATA ANALYTICS IN MACHINE LEARNING FOR INTERNET OF THINGS

R. Priyanka
Department of Computer Science and Engineering, Coimbatore Institute of Engineering and Technology, India

Abstract

The Internet of Things (IoT) is rapidly merging sensor and actuator technology, embedded systems, wireless communication, and data analytics. Massive volumes of data are being generated by sensors in IoT systems across a wide range of businesses. If IoT can be used to analyse data to improve decision-making, productivity, accuracy, and income, businesses and life-enhancement paradigms can profit from IoT. We need a new paradigm for obtaining knowledge from IoT data to improve quality of life. Smarter IoT requires deep learning since these services are the cornerstone for IoT applications. Deep learning IoT data analytics research is vital in this setting. Here, we discuss deep learning architectures, their relevance in IoT data analytics, and prospective use cases. There is a discussion of current research issues and potential future research initiatives.

Keywords:

Data Analytics, Machine Learning, IoTs

1. INTRODUCTION

The Internet of Things (IoT) paradigm is both revolutionary and a facilitator of automated and convenient lifestyles for modern people. IoT development can be credited to developments in computation, networking, and application design during the previous decade [1]. As a result of the IoT, the entire human species is now under its zone of influence.

In our daily lives, we utilise smartphones, home assistants like Google Play, smart vehicles, building automation systems, and unmanned aerial vehicles such as drones for environmental monitoring and recreational purposes to make our lives easier [2].

The massive growth of Internet of Things (IoT) devices extends beyond the devices themselves to the widely dispersed storage centres, such as the back-end cloud services. For further analysis, since so much data is created by IoT devices and their supporting platforms, it is sent and processed at backend cloud storage centres [3] for storage and further analysis.

Continuous streams of raw data generated by IoT devices cannot be converted into meaningful knowledge without the use of techniques like knowledge discovery and artificial intelligence. Due to the interconnected nature of the Internet of Things (IoT), e-agriculture and e-health, as well as smart electrical grids and smart vehicles, all generate different types of heterogeneous data due to the Internet of Things (IoT) [4].

Custom protocols for IoT devices take into account their resource limitations in order to preserve the power consumption associated with device activities. Many IoT application-layer protocols exist, the most prominent of which are: Constrained Application Protocol (CoAP); Message Queuing Telemetry Transfer (MQTT); Advanced Message Queuing Protocol (AMQP); and HTTP [5].

2. PROBLEM DEFINITION

These protocols can only be used with IoT devices that have either a constant power supply or a renewable power source. Longer messages are easier to send using these protocols, which consume a lot of electricity. The CoAP protocol, on the other hand, is lightweight and optimised for IoT devices with limited computational power and network bandwidth, as opposed to other protocols. An IoT device with a greater level of processing power, connectivity, and storage capacity should use the HTTP protocol, which is the most resource-intensive. The amount of data generated by Internet of Things (IoT) devices is so huge that it can only be processed in a limited way on the device itself before being sent to a central computing node or a cloud storage facility for additional processing or analysis [6]. In the context of machine learning, the process of generating models from training data is automated, meaning that little to no human participation is required during the process. It is therefore possible to totally automate the process of classifying data [7]. It impossible to overstate the importance of data analytics for IoT data processing, and machine learning is a key contributor to speeding up the processing of enormous volumes of data generated by IoT devices.

In the modern era of computation and data storage, cloud, fog, and edge computing are among the most popular options. Incorporating these concepts with the Internet of Things (IoT) results in a comprehensive framework for data collection and analysis. Using such a framework, machine learning techniques may be applied to IoT data in real time, allowing for sophisticated data analytics for the IoT. IoT data processing and analysis are made possible by the cloud paradigm, a centralised data storage model that provides multiple services such as software-as-a-service (SaaS), platform-as-a-service (PaaS), and infrastructure-as-a-service (IaaS) [8]. Edge computing, on the other hand, allows IoT data to be processed and analysed on localised computer nodes such as base stations, which are located closer to the IoT network. Consequently, the costs associated with moving IoT data to centralised nodes are minimised. It is possible to collect and analyse IoT data without storing it at the network edge or even in a centralised storage facility using fog computing, which is a solution that sits somewhere in the midst of cloud computing and edge computing. A virtual platform for processing and analysing IoT data is introduced by the fog paradigm rather than a physical one only found at the network edge [9].

3. IOT-ML CONVERGENCE

For resource-limited IoT devices, the convergence of machine learning and IoT could lead to a significant increase in efficiency, accuracy, productivity, and overall cost savings. By combining machine learning algorithms and the Internet of Things, we can increase communication and computation performance, controllability, and decision making. The Internet of Things (IoT) holds immense promise for enhancing human well-being and fostering economic development because of its vast array of ubiquitous sensing devices and enhanced communication capabilities (toward Industry 4.0). Machine learning and artificial intelligence have greatly enhanced the possibilities of the Internet of Things (IoT). Advanced machine intelligence algorithms have made it feasible to mine the enormous volume of IoT sensory data in order to better understand a wide range of real-world situations and make key operational decisions. Machine learning and the Internet of Things (IoT) must work together to solve complex real-world issues and meet computational and communication constraints. For the following reasons, IoT data analytics has recently garnered a lot of attention and importance:

Distributed IoT devices produce a large amount of data. According to Ericsson mobility research, there will be 18 billion linked IoT devices worldwide by 2022. The number of essential applications where IoT devices are used will continue to grow over time because of the widespread usage of IoT devices. By mining this enormous amount of data rapidly and intelligently, intelligent data analytics will play a key role in identifying and predicting the future states of any process or system.

a wide range of data kinds from a variety of different sources. Mobile phones, PC/laptop/tablets, and short-range and long-range Internet of Things (IoT) devices are just some of the many types of IoT devices available. The characteristics, forms, and properties of the data vary because of the heterogeneity of the data. In addition, the data sources vary according to the various IoT application domains. Medical IoT devices, for example, will be distinct from smart home IoT devices. Due to its heterogeneity, it has become increasingly difficult to maintain the quality, processing, and storage of data. When it comes to data sources, the authors in [4] raise a number of important problems. Sample procedure for high-frequency streaming data; how to deal with noise cancellation and filtering; how to merge heterogeneous data sources; how to create situation awareness and knowledge; how to gather and store heterogeneous data sources to meet the application constraints; and how to create situation awareness and knowledge. [5].

There is uncertainty in the data streams of the Internet of Things. Practical data analysis involves a great deal of uncertainty [8]. During data transfer, any IoT device or communication channel may fail. This might occur in the data stream. IoT data streams are riddled with errors and omissions, necessitating the use of sophisticated analytics to clean them up. It is possible that even a cyber-attack could lead to a lack of confidence in data. Accurate evaluation, propagation, and representation of uncertainties, as well as models and strategies to address these issues, are essential for improving decision accuracy [9].

Scalability and effectiveness must be carefully balanced. The cloud is where the majority of IoT data analysis is done. Transferring data from an IoT device to the cloud is time-consuming and expensive, making it difficult for time-sensitive applications to scale. A significant number of cars may be required to make choices in real-time or near real-time in a connected vehicle environment. When the number of cars increases, it critical to strike a balance between speed and accuracy.

4. PROPOSED METHOD

Various data analytics methods are discussed in this section. The Fig.1 depicts the major components of analytical classes.

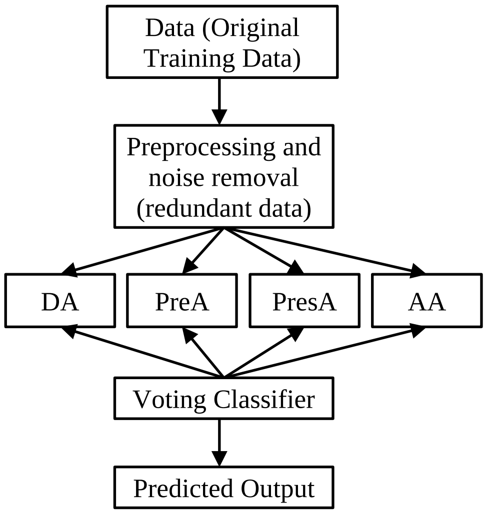

Fig.1. Data Analytics Model

4.1 DESCRIPTIVE ANALYTICS

From a few smart devices, IoT systems can collect data and communicate it to a cloud environment. Using modern machine learning algorithms, it is always possible to get insights into the past based on previous data. Descriptive analytics is a branch of research that uses machine learning-based algorithms to process and summarise raw data and deliver actionable insights. In descriptive analytics, data aggregation, data summarization, mathematical logical processes, and so on are some instances of data mining (e.g., clustering algorithms). A large amount of data is needed for descriptive analytics. Using high-performance computers and IoT cloud analytics, recent technological advancements have shown that cloud storage can store enormous volumes of IoT data.

4.2 PREDICTIVE ANALYTICS FOR IOT

It is possible to forecast future trends or patterns in data using predictive analytics, which relies on past data and advanced statistical or machine learning approaches. To sum it up, it makes predictions about the future based on patterns and correlations found in previously collected data. For a variety of purposes, predictive analytics has been widely employed, such as in predictive maintenance, price forecasting, supply-demand forecasting, and the prediction of any result. SAS, a renowned analytics company, says there are two types of predictive models: classification-based models, which conduct analysis by class membership, and regression-based models, which forecast a number based on past observations and likelihood [80]. Statistical regression-based models, decision trees, and neural network or deep neural network-based models are among the most advanced predictive modelling techniques. Bayesian analysis, gradient boosting, ensemble model-based analysis, and so on are all popular techniques. These predictive analytics methods rely heavily on data for making decisions. It is possible to use the Internet of Things (IoT) paradigm to enable the collection of data from smart IoT devices and to create a framework for analysis using the cloud or edge of the network.

4.3 PRESCRIPTIVE ANALYTICS FOR IOT

Prescriptive analytics uses data analysis to advise on actions to be taken in the future. This type of research not only forecasts the future but also makes recommendations for how that future state should be achieved. Descriptive and predictive analytics are used in tandem to create a future scenario analysis. Predictive analytics tells us when and where an event will take place based on past trends, but prescriptive analytics goes a step further by analysing the potential consequences of those forecasts. Prescriptive analytics is commonly used to improve company outcomes. Using cloud/edge computing, big data analytics, and machine learning to make business intelligence-based decisions is a good fit for predictive analytics in an Industrial IoT (IIoT) configuration. The deployment of business intelligence tools and analytics within an IoT-cloud platform can assist in making optimal decisions.

4.4 ADAPTIVE ANALYTICS FOR IOT

During deployment, real-time data must be incorporated into the predictive analytics results. Adaptive analytics is used to alter or optimise the process outcome depending on recent data history and correlations. Improved model performance and fewer mistakes can be achieved by conducting this type of study. Adaptive analytics has the advantage of being able to alter its results when fresh data is obtained. Adaptive analytics is ideal for real-time stream data processing in an IoT scenario. Adaptive analytics can also be used for real-time analysis of dynamic data streams, such as those seen in malware.

5. VALIDATION

IoT areas such as smart homes, farming, smart cities, and health can all benefit from the suggested intelligent framework ability to exchange and generate knowledge autonomously. It is important to note that the framework does not address how IoT devices might adapt to restricted computation and network resources (as described by Model 1 in Fig.2). Rather than trying to improve network and compute performance, we demonstrated how adaptive learning can be done inside current IoT models. Thus, the architecture takes advantage of the fact that learning, or translating data into knowledge, can be done either in the cloud (as in Model 2) or at the edge (as in Model 3).

5.1 LATENCY

The framework can leverage ontologies to integrate IoT devices with intelligent systems by adopting Models 2 and 3 as shown in Figure 3. Evaluating the importance of the application is an important step in making a decision on which model to use. However, if the network latency and network availability requirements for a remote, mission-critical application have been optimised, one can choose to use Model 2 (i.e., cloud-based resources). One of the benefits of this paradigm is that the cloud-based SO can learn from remote SOs in the background, regardless of the network delay that connects them to the remote app. The drawback is that if the demand for real-time choices grows, network latency may become a bottleneck. Another option is to go with Model 3. (i.e., having all resources close to the consumer). In this approach, the SO can provide consumers with real-time knowledge, but it relies on the communications network capacity and availability to start knowledge exchange with remote SOs.

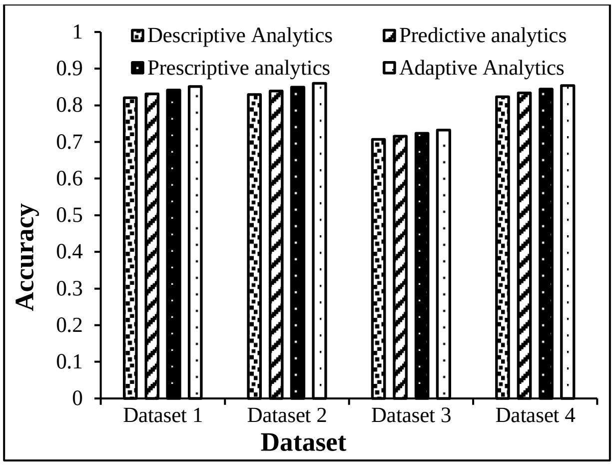

Fig.2. Accuracy of Analytics

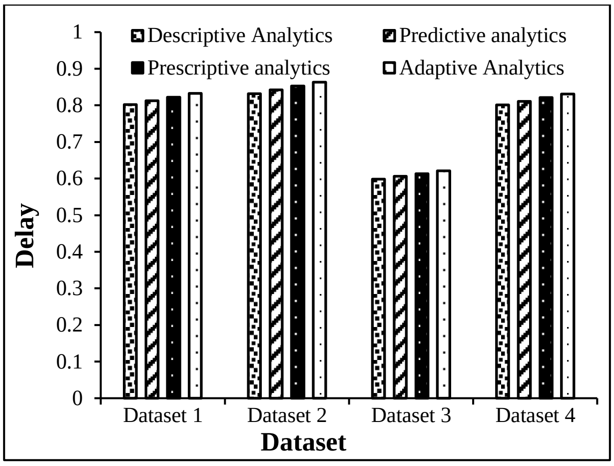

Fig.2. Latency of Analytics

6. CONCLUSION

There is an enormous amount of data generated by sensors connected to the IoT. Using the IoT to analyse data and develop business and life-enhancement paradigms will be beneficial if this can be done. Using IoT data to improve quality of life will require a new paradigm for extracting hidden information and inferences. A smarter IoT relies on the development of Deep Learning services, which are the building blocks of IoT applications. In this setting, research in deep learning IoT data analytics is essential. On the subject of IoT data analytics, this study focuses on Deep Learning architecture. In this work, the current research issues and future research directions are addressed.

REFERENCES

[1] Luca Catarinucci et al., “An IoT-Aware Architecture for Smart Healthcare Systems”, *IEEE Internet of Things Journal*, Vol. 2, No. 6, pp. 515-526, 2015.

[2] Sarfraz Fayaz Khan, "Health Care Monitoring System in Internet of Things (loT) by Using RFID", *Proceedings of 6th International Conference on Industrial Technology and Management*, pp. 198-204, 2017.

[3] Prosanta Gope and Tzonelih Hwang, "BSN-Care: A Secure IoT-Based Modern Healthcare System using Body Sensor Network", *IEEE Sensors Journal*, Vol. 16, No. 5, pp. 1368-1376, 2016.

[4] Byung Mun Lee, "Registration Protocol for Health IoT Platform to Share the Use of Medical Devices", *International Journal of Bio-Science and Bio-Technology*, Vol. 7, No. 4, pp. 1-10, 2015.

[5] R. Shantha Mary Joshitta and L. Arockiam, "Key Generation Algorithm using Soft Set for Data Security in Internet of Things", *Proceedings of 3rd International Conference on Internet of Things*, pp. 367-372, 2018.

[6] R. Shantha Mary Joshitta and L. Arockiam, "Security in IoT Environment: A Survey", *International Journal of Information Technology and Mechanical Engineering*, Vol. 2, No. 7, pp. 1-8, 2016.

[7] A. Cristina Enache and Victor Valeriu Patriciu, "Intrusions Detection Based on Support Vector Machine Optimized with Swarm Intelligence", *Proceedings of IEEE International Symposium on Applied Computational Intelligence and Informatics*, pp. 153-158, 2014.

[8] M. Suhasini and B. Srinivasu, "Emotion Detection Framework for Twitter Data using Supervised Classifiers", *International Journal of Computer Applications*, Vol. 5, No. 3, pp. 1-12, 2020.

[9] Minal Deshpande and Preeti Bajaj, "Performance Improvement of Traffic Flow Prediction using Combination of Support Vector Machine and Rough Set", *International Journal of Computer Applications*, Vol. 163, No. 2, pp. 31-35, 2017.

[10] Xin-She Yang, "Engineering Optimizations Via Nature Inspired Virtual Bee Algorithms", *Artificial Intelligence and Knowledge Engineering Applications: A Bioinspired Approach*, Vol. 3562, pp. 317-323, 2005.