A Survey on Data Stream Mining Towards the Internet of Things Application

Conference Paper · December 2019

DOI: 10.1109/TIMES-iCON47539.2019.9024597

6 authors, including:

Saifun Nahar
University of Missouri - St. Louis
16 PUBLICATIONS **160** CITATIONS
SEE PROFILE

Happy Nkanta Monday
Oxford Brookes College of Chengdu University of Technology
50 PUBLICATIONS **574** CITATIONS
SEE PROFILE

Grace U. Nneji
Oxford Brookes College of Chengdu University of Technology
48 PUBLICATIONS **447** CITATIONS
SEE PROFILE

Hassan Sani Abubakar
University of Electronic Science and Technology of China
16 PUBLICATIONS **79** CITATIONS
SEE PROFILE

A Survey on Data Stream Mining Towards the Internet of Things Application

* Saifun Nahar[1], Ting Zhong[1], Happy N. Monday[2], Michael O. Mills[2] , Grace U. Nneji[1], Hassan S. Abubakar[3]
[1]School of Information and Software Engineering
[2]School of Computer Science and Engineering
[3]School of Information and Communication Engineering
University of Electronic Science and Technology of China
Chengdu, Sichuan P.R China
* sainahbd@163.com

***Abstract*—In the era of the Internet of Things, a widespread of applications depend on time with the various number of different data generated and collected from different devices available. These devices depend on the type of application. These fast stream data are real-time and large in dimension for the purpose of making decision as well as predicting future occurrence and analytics. Datastream analytics of internet technology for both businesses and everyday life is very valuable in terms of developing good quality of life. In this study, first of all, we focus on the concept of Internet of Things and its relationship with its architecture, large and flowing data. In addition, the approach of Internet of Things applied knowledge discovery process and deep learning frameworks are presented in this paper. Finally, the Internet of Things and its features are introduced in this work as well as the commonly used tools.**

***Keywords—internet of Things; Big Data Analytics; Deep Learning; Flowing Data Mining; Data Processing Platforms*.**

I. INTRODUCTION

Internet of Things can be viewed in a unique way where objects establish communications with themselves in a common worldwide network. Objects in this network interact with each other using a specific communication protocol or various communications protocols [1] [2]. According to research, 10-11 billion devices are connected to the internet presently. It is predicted that the number of gadgets connected on the internet will increase to 50 billion by 2020. According to the same research, in 2003 the ratio of interconnected gadget to a person was 0.08. The 2020 estimate is 6.48. Also in 2020, the traffic information generated by the typical household appliance produced in 2008 was 50% bringing to a huge trillion GB of internet traffic in approximation [3]. Internet of Things exposes information hidden in its large data so as to improve the quality of lives by removing complex task [4]. To subdue the difficult and complex task associated with the traditional methods, new inferences and learning approaches technologies, algorithms, infrastructures are required. Fortunately, fast data processing and more developed machine learning techniques improvements enable large data analytics and information extraction [5]. Beyond big data analysis, Internet of Things data with high-speed data streams and time accuracy actions developed to support applications with an analytical class, fast and flowing data analytics concept was formed. Some researchers have found cloud infrastructures and service analytics frameworks and approaches. Unnecessary communication closes the data source to avoid delays. [6] [7]. In this study, the importance of data analysis in the Internet of Things in terms of the techniques and tools used are summarized. Deep learning application is also introduced in the rest of the study.

II. THE INTERNET OF THINGS

Wireless communication is becoming increasingly widespread. This the simplest idea of the concept, specific addressing RFID, sensors, electronic labels, etc. Around us, different objects interact with each other and also collaborate for common goals [3]. The Internet of Things structure is attached to each other via the Internet. A small group of objects can connect with other objects. Therefore the architecture must be flexible. The simplest form of studies done on the Internet of Things is; detection layer, network layer, and application layer. It forms the new platform for the Internet of Things over the years. Layer views are specified an architecture widely used is five layers [3];

- Sensors and actuators in the Object Detection Layer obtained by different devices via information such as temperature, humidity, weight, speed, acceleration, and location brought together. Configure different types of objects. The plug and play mechanism is used in this layer.

- From the Detection Layer of the Object, Abstraction Layer secures the Service. Management Layer of the resulting data communications via channels.

- Service Management Layer. Address and name of the service is the layer where it is paired with requests. Internet of Things program is written to support an application with different specific hardware platforms interacting with objects of different types.

978-1-7281-3755-1/19/$31.00 ©2019 IEEE

- Application Layer meets the need of the object. Internet of Things system created for solutions to meet the layer's needs. This layer provides information to request such as the amount of moisture in the air, temperature.

- The Business Layer is designed to manage the management layer. Application with the data obtained from the layer of the business model is the creation of graphs, tables of this layer responsibility.

A. Big Data and Data Flow in the Internet of Things: Data volume, data types, variety, and data rate are the features of big data [8], [9]. For the collected data, features such as diversity, speed, data volume, reality, and variability can be counted as value. Internet of Things flowing data from traditional big data separating properties are as follows [10]-[12];

- Numerous amounts of data received from devices are distributed and continue flowing data. Applications are in real-time and updated periodically.
- Internet of Things applications are mostly sensors. For this reason, location and timestamp are available.
- Due to the size of data in the Internet of Things applications, there are data errors and noise during transmission and transmission of data.
- Data streams are self-directed amongst objects and occur in a stochastic process.
- For correct processing of data, transmission order is quite important.
- System data elements, data flow or data edit around the streams have corresponding control.
- Once an element is processed from the data stream, it is either removed or archived. Most especially, the memory cannot be restored if not stored.
- To query the data flow in a traditional relational model is quite different
- Unlike stack data, data must be processed.

III. Information Discovered On The Internet

Getting a large amount of data on the internet is called big data. To keep this data while analyzing, revealing hidden patterns becomes difficult. Objects Information discovery process of data obtained in the internet environment with the implementation of interesting and unknown information using traditional machine learning or deep learning method with learning characteristics [4]. Some studies show how well machine learning algorithms have been applied to data stream mining. However, it is a topic of intensive study. There are many proposed approaches and algorithms [13]. Deep learning algorithms have been proposed by many researchers in data stream mining due to its learning new features and use it for a classification task. These techniques are considered problem-oriented but may not be the same depending on the approaches and nature of data [6].

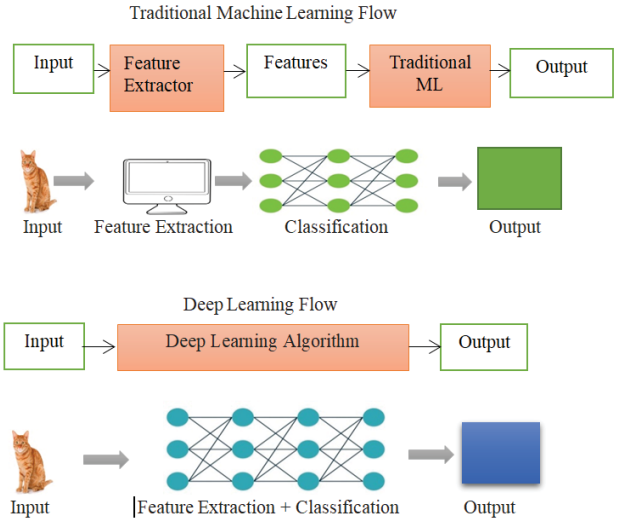

Fig 1: Machine Learning VS Deep learning

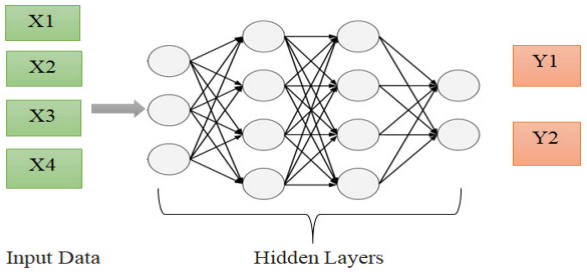

Fig 2: Deep Learning Architecture

Deep learning is self-learning architecture of neural network [23]. A deep learning algorithm is capable of learning features and extraction. Each successive layer takes the output from the preceding layer as input [14] [15]. Different frameworks are available for the use of deep learning architectures in different areas. Each of these frameworks supports deep learning architecture, and optimization algorithms; depending on the development, ease of use and advantages [16].

Table 1: Properties of Deep Learning Frames

Tools	Core Language	Interface	Benefits	Drawbacks
H2O	Java	R, Python, Scala, REST API	Wide Range Of Interface	A limited number of models are supported. It is not flexible.
Tensor flow	C++	Python, Java, C, C++, Go	Long Short Term Memory(LSTM) is quite fast. Support Visualization Network.	More Python-based frameworks are slowly trained.
Theano	Python	Python	Support various model. On the GPU(Graphics Processing Unit) LSTM is quite fast.	Many low-level Applications It has Program Interface.
Torch	Luna	C, C++	Support various model. Good documentation. The debugger has debugged messages.	It is open to learning a new language.
Caffe	C++	Python, MATLAB	Provides a collection of reference models. Easy to change platform. Has effective results in folded networks	It is not very effective in iterative networks.
Neon	Python	Python	It has a fast training time. Easy platform adaptation and supports modern architectures	CPU (Central Processing Unit) multiple uses.
Chainer	Python	Python	Supports modern architectures. Complex architectures can be easily placed. The dynamic model can be changed.	Slower forward in some scenarios calculation feature.

IV. Data Analytics Platforms

In accordance with the needs of organizations, Internet of Things plays a huge role in real-time data collection for analytics platforms. The number of data produced has doubled in size due to increased usage of devices such as phones, sensors, etc. Data analytics platform features [18] summarized in the study;

- Analyze flowing data in real-time and should be capable of reporting.
- Users regardless of where they are located should be able to login to the platform.
- Independent usage of interfaces should be designed as easy.
- Be able to perform various types of data and data should be easy to adapt.

TABLE 2: Features of Flowing Data Analytic Tools

Flowing Data Analytics tools	Features
Stream Analytics	Integrates many machines into one platform. Infrastructure technology does not require much user attention
Information	Provides off-pack services. Keeps all data connected and eliminates the hassle of manual code writing.
SAP Event stream processor	Displays future event flows using trends, patterns, and a correlation between them. Uses notifications and alarms to keep you informed of opportunities and threats.
Oracle Stream Analytics	Includes interrelated Visual Geoprocessing and Geoforce for spatial analytics. New Expressive Patterns with machine learning capabilities for spatial, statistical, anomaly detection library.
SQL Stream	It is a distributed platform developed according to open standards. Flexible development, scalability, analytical operations and reuse of applications through the application program interface features.
Apache Flink	It has limited and unlimited data processing capacity. Run applications in any environment and scale. Power transfer feature is available in the memory performance. It ensures accurate results even in irregular, late data. High tolerance to error. It can work efficiently on thousands of nodes and manage latency. In addition to data-driven windowing, it has flexible windowing capability based on numbers and sessions. Applications can update or process historical data without any loss and interruption. The application program interface is very useful. User-friendly and developed to cover all common jobs. Configuration does not require the use of memory, network, serializer.
Spark	Hadoop can work on Mesos, in the cloud or alone. Different distributed data such as HDFS (**Hadoop Distributed File System**), Cassandra, HBase, S3 resources. It empowers different library stacks such as Mlib, Dataframes, SQL, GraphX and Spark Streaming.

V. Mode of Work

Data are generated from meaningful yet different sources for the application of IoT technology. Internet of Things is playing a major role in today's world. Its impact and technology vary across different applications such as vehicle tracking, parking systems, payment system and utility management system with sensors designed to emit rays. These sensors are cameras that sense motion, sound as well as an emitting ray. The pressure sensor is sensor cameras that emits infrared rays to sense pressure [24]. These sensor utilizes similar technology by sending short messages as SMS and Necessary warnings to users [19]. With the progress of wearable technologies, the monitoring of the health parameters of individuals has facilitated wearable sensors capable of reading blood pressure, heartbeats, and sleeping patterns. Based on this technology, the data obtained can be used to prescribe a drug to patients with having to schedule a physical meeting. This technology can also prevent patients from taking the wrong dose of drugs at the wrong time [20]. Visually Impaired Navigation System is designed for people with disabilities to help them while shopping. The supermarket is divided into cells radio placed in frequency tags and the navigating tags are mapped [21]. Individuals parsing the sound of the smartphone and using the recognition system to go to the phone notifies and go the person with Bluetooth and WLAN technologies routing for the hearing impaired temperature sensor taken with temperature sensor reached through a status check to a control center. Disable individual with shining light or vibration warning is provided [20] [22]. Smart environment Sensors placed in homes and workplaces to measure ambient temperature, humidity, and light for personal preference. These parameters can be adjusted according to weather conditions. Data obtained by working on peoples' personal habits adjustments are possible [3]. Smart city applications for garbage separation, traffic control, energy network as well as forecasting the air quality [16] [24].

VI. Conclusion

Internet of Things, deep learning and flowing data mining concepts have contributed positively to our lives, society, and the world as a whole. Big data research has become quite popular amongst researchers nowadays. Internet of Things technology through the advent of big data has changed the way we perceive information. Stream data mining offers a platform for real-time data analytics. The impact of the application of data stream mining towards IoT can not be overemphasized due to its numerous advantages. Data obtained are used for decision making and prediction of events. Finally, This paper has presented to its readers the numerous benefits and application of data stream mining towards IoT.

Acknowledgment

Work supported by University of Electronic Science and Technology of China.

References

[1] C.W. Tsai, C.F. Lai, M.C. Chiang, L. T. Yang, "Data Mining for the Internet of Things: A Survey." IEEE Communications Surveys and Tutorials, vol. 16, pp.77-97, November 2013.

[2] M. A. Alsheikh, S. Lin, D. Niyato, and H.P. Tan, "Machine Learning in Wireless Sensor Networks: Algorithms, Strategies, and Applications," IEEE Communications Surveys & Tutorials, vol. 16, pp.1996-2018, April 2014.

[3] M. Mazhar Rathore, Anand Paul, Won-Hwa Hong, Hyun Cheol Seo, Imtiaz Awan, Sharjil Saeed, "Exploiting IoT and Big Data Analytics", Sustainable Cities and Societies, vol.40, pp.600-610, July 2018.

[4] Atzori L., Lera A., "Internet of Things: Survey", Computer Networks, vol.54, pp.2787-2805, June 2010

[5] M. Zaharia, M. Chowdhury, T. Das, A. Dave, J. Ma, M. Mccauley, M. Franklin, S. Shenker, and I. Stoica, "Fast and Interactive Analytics Over Hadoop Data with Spark", USENIX Login, vol.37, no.4, August 2012.

[6] C. Engle, A. Lupher, R. Xin, M. Zaharia, M. J. Franklin, S. Shenker, and I. Stoica, "Shark: Fast Data Analysis using Coarse-grained Distributed Memory", Proceedings of 2012

ACM SIGMOD International Conference on Management of Data. ACM, 2012.

[7] M. Chen, S. Mao, Y. Zhang, and V. C. Leung, "Big Data: Related Technologies, Challenges, and Future Prospects", Springer, pp.19-32, July 2015.

[8] W. Fan, A. Bifet, "Mining Big Data: Current Status and Forecast to The Future", ACM sIGKDD Explorations Newsletter, vol.14, pp.1-5, December 2012.

[9] H. Hu, Y. Wen, T.S. Chua, and X. Li, "Toward Scalable Systems for Big Data Analytics: A Technology Tutorial", IEEE Access, vol.2, p.652-687, June 2014.

[10] Y. Demchenko, P. Grosso, C. De Laat, and P. Membrey, "Addressing Big Data Issues in Scientific Data Infrastructure", International Conference on Collaboration Technologies and Systems (CTS), July 2013.

[11] Sorour S., Mohammedi M., "Deep Learning for IOT Big Data and Streaming Analytics: A Survey", IEEE Communications Surveys & Tutorials, vol.20, pp.2923-2960 June 2018.

[12] Guoqiang Z., Li-Na W., Ling X., Dong J., "An Overview on Data Representation Learning: From Traditional Feature Learning to Recent Deep Learning", The Journal of Finance and Data Science, vol.2, pp.265-278, December 2016.

[13] Z. Fadlullah, F. Tang, B. Mao, N. Kato, O. Akashi, T. Inoue, and K. Mizutani, "State-of-the-art Deep Learning

[14] S. Han, J. Pool, J. Tran, and W. Dally, "Learning Both Weights and Connections for Efficient Neural Network" in

Advances in Neural Information Processing Systems, vol.1, pp.1135-1143, December 2015.

[15] Xiaofeng X., Siping L., "IoT Data Analytics Using Deep Learning", IEEE, 2016.

[16] Ching Yu Chen, Jui Hsi Fu, Today Sung, "Complex Event

Processing for the Internet of Things and Its Applications", IEEE, vol.20, pp.2923-2960, August 2014.

[17] Top 20 Free Open Source and Premium Stream Analytics Platforms. (2018, 1 Eylül)

[18] C.W. Tsai, C.F. Lai, M.C. Chiang, and L. T. Yang, "Data Mining for the Internet of Things: A Survey", IEEE Communications Surveys & Tutorials, vol.16, pp.77-97, November 2013.

[19] Y. J. Fan, Y. H. Yin, L. Da Xu, Y. Zeng, and F. Wu, "IOT Based Smart Rehabilitation System", IEEE Transactions on Industrial Informatics, vol.10, pp.1568-1577, January 2014.

[20] M. Price, J. Glass, and A. Chandrakasan, “A Scalable Speech Recognizer with Deep-Neural-Network Acoustic Models and Voice-Activated Power Gating”, Proceedings of the IEEE ISSCC2017, March 2017.

[21] M. C. Domingo, "An Overview of The Internet of Things for People with Disabilities,", Journal of Network and Computer Applications, vol.35, pp.584-596, March 2012.

[22] N. D. Lane, S. Bhattacharya, P. Georgiev, C. Forlivesi, and F. Kawsar, “An Early Resource Characterization of Deep Learning on Wearables, Smartphones and Internet of Things Devices”, Proceedings of the 2015 International Workshop on Internet of Things towards Applications, pp.7-12, November 2015.

[23] I. Kok, M.U. Simsek, S. Ozdemir, “A Deep Learning Model for Air Quality Prediction in Smart Cities”, IEEE International Conference on Big Data, December 2017.

[24] S.E. Bibri, “The IoT for Smart Sustainable Cities of The Future: An Analytical Framework for Sensor-based Big Data Applications for Environmental Sustainability”, Sustainable Cities and Society, vol.38, pp.253, April 2018.

.