IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Exploratory Study for Big Data Visualization in the Internet of Things

Sudhir Allam, *Sr. Data Engineer, Department of Information Technology, USA*

Abstract—*While the Internet of Things (IoT) gains momentum in the tech industry, large amounts of wireless devices are developed, tracking a vast variety of infrastructure, which constantly optimizes data in different fields such as medical, transport and logistics, energy, agriculture intelligence, building automation, and data producing industries [1]. Big data techniques play a major role in IoT processes as instruments for l visualization to produce useful information in real-time and enable strategic decision-making. This paper includes an extensive analysis of how big data visualization is beneficial to the IoT approaches, software, and techniques. By examining the visual analytics, the paper places the data visualization within the visual analysis phase. It offers a study of different tools available for data visualization and analyzes guidelines for each of them, considering the specific circumstances of the individual use case. Although the big data methods are separated from each IoT area, the paper will explore visualization challenges and how big data shapes the IoT. To fully explore the topic, the paper requires literature reviews to establish its composition. While this paper does not show any findings, it provides a summary of what has been achieved to date in big data visualization in IoT and in-depth learning application in this field of science. The role of big data on IoT visualization will also be presented. It emphasizes illustrating the main concepts of Big Data visualization in real-time.*

Keywords: *Big Data, IoT-big data visualization, Data visualization, Internet of Things, Machine learning, Visual analytics*

I. INTRODUCTION

Data visualization is a pictorial or graphical display of data. It helps decision-makers to visually interpret analytics to understand difficult principles or to detect emerging trends. By utilizing technologies of interactive simulation, they will take the idea a step forward using diagrams and graphs, interactively changing which data they display and how it is stored [1]. The Internet of Things (IoT) has been one of the most important and evolving innovations that enhance the quality of life. IoT integrates a wide variety of heterogeneous tools to efficiently collect diverse forms of real-world data. To better the everyday life of people, IoT data is utilized to collect valuable knowledge, through context-aware technologies. Since data is usually accompanied by contextual information (time, place, status, etc.), IoT becomes a useful and abundant source of contextual data with variety (multiple sources), speed (real-time compilation), veracity (data uncertainty), and reliability [1, 2]. The collaboration between Big Data and IoT has led to the implementation of intelligent services for several diverse infrastructures. While IoT evolves quickly, big data applications play a vital part in promoting critical decision-making as visual analysis instruments that provide useful real-time information in IoT infrastructure. Big IoT systems use a wide range of sensors, leading to a very large volume of data gathered [2]. Two activities are important in the sense of IoT data analysis: searching vast volumes of data to identify subsets and trends of interest, and reviewing the existing data to create assessments and forecasts. This study offers detailed illustrations to gain insight into IoT data. Visual analytics is a research methodology that employs data mining, measurements, and visualization to aid in the exploration of large volumes of data. Interactive visualization applications integrate automated processing with human interfaces that enable users to monitor the data analysis process to produce useful insights into decision making [2]. They have personalized data visualization approaches to support the user to engage with them to display and report on interesting facts from various viewpoints. Methods in data analytics use deep learning and AI methods to automatically extract trends and forecast. Owing to their black-box functionality, AI approaches are generally untrustworthy to their users, which gives little insight into their findings. Visual analysis may be used to clear and clarify big data approaches by analyzing both their findings and their function.

Data visualization and analytics are also the key elements of business analytics and convert the proliferation of big data into relevant insights across digital frameworks. The Big Data age has indeed recognized the existence of massive, complex, messy, heterogeneous datasets. It is still more burdensome for many people to convert a data-curious individual into someone who can view and interpret the data with little to no assistance and experience [3]. Data researchers also needed more time and resources to overcome popular issues including overpotting while viewing IoT/big data. This analysis aims to explore how the visualization data can be significant for IoT, its challenges, relevant work, and feasibility to create a basic data visualization system for IoT systems.

II. RESEARCH PROBLEM

The main research problem that this paper aims at solving is how big data visualization can help resolve challenges associated with the successful deployment of the Internet of Things. There have been issues in most technological processes when managing and deploying IoT due to the ever-increasing data consumption devices have been increasing at an exponential rate creating challenges when it comes to data processing, the efficiency of collecting data, analysis, and maintaining the security of such devices. To address these concerns, this paper will look at how big data visualization can converge with IoT to manage and create opportunities from the IoT systems. Data can be found anywhere! First, from the user's website, all of the data and behavioral patterns are monitored, before they sign out. Stats from smart sensors that track the fitness for the massive logistics and freight business that controls the stock of the supply chain, resource monitoring all results in a lot of big data. All this information is worthless unless they are proper utilized for a certain reason [4]. The analysis of data and the prediction of potential trends as well as where to emphasize them are some of the important advantages. This intelligent data visualization software and technology allow data to be sliced and diced down to the smallest detail [5].

III. LITERATURE REVIEW

A. Why the Internet of Things forms big data (IoTs)

Big Data is an all-around concept for a wide variety of data that comprises a vast amount of heterogeneous data, organized, semi-structured and unstructured. Big data and IoT have a strong collaborative relationship, even though an AI framework is introduced to process and take decisions, a powerful ecosystem is built. The more IoT devices linked or more sophisticated AI systems the larger the importance of big data hardware, as data storage is the archive and data center [5,6]. To determine what is required, which emphasizes investing in productive hardware intelligently, or maximizing infrastructure architecture, quality, and processing, it largely relies on the power of large data hardware.

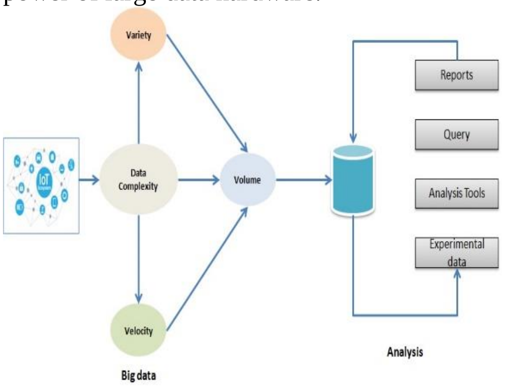

Fig i: Relationship between big data and IoT

B. Big data analytical role in IoT

Smart devices are vital components of the Internet of Things; these devices produce a vast volume of data that must be processed and examined in real-time. Predictive and large data modeling are at stake here. Besides, large data processing systems use IoT to make it easier to work but still reveal certain challenges. Big data can be used in IoT thanks to the vast rollout of sensors and Internet applications. Besides, data analysis in big data poses difficulties owing to the short IoT system end computing, networking, and storage media [6]. The function of big data in IoT becomes crucial as the whole IoT framework serves as a data-generating source. Big data analytics is an emerging method for evaluating data transmitted over the network in IoT, which aids in leading the way to enhance decision-making. The Big data process can accommodate a vast volume of data that is processed in real-time and maintained using various computing strategies such as Microsoft Azure [6].

C. *Challenges in IoT with Big Data Analytics*

In this segment, the e key concerns of big data analysis in IoT: swift development in different applications within IoT will be addressed. These challenges are explained below:

1. **Visualization of data:** Commonly, heterogeneous data are constantly produced. i.e. organized, unstructured, and semi-structured in numerous formats, rendering it impossible to explicitly display these files. Data are needed to help visualize and interpret industrial decisions accurately and in good time and to improve the productivity of the sector [7]. Types of Data Visualization in Business Analytics is indeed something one should learn more.

2. **Privacy and confidentiality:** Any smart device in a globally linked network represents an IoT device, and pays more regard to privacy and knowledge disclosure, whether it is used by humans or computers. This critical data can also be kept private and secure because the data generated includes users' details.

3. **Data storage and management:** As the volume of data collected by internet-connected devices grows at an ever-increasing pace, and the storage space of Big data systems becomes small, storing and managing such a large number of data becomes a major challenge [7]. Some processes and systems must be designed to collect, save and manage this data.

4. **Integrity:** connected computers are capable of detecting, interacting, exchanging, and performing analyzes for various purposes. These devices ensure the consumer cannot exchange their data indefinitely, and data assembly methods must effectively implement scale and integrity requirements for some basic procedures and rules.

5. **Power captivity:** For the seamless and uninterrupted running of IoT operations, Internet-enabled computers should be connected to an infinite power supply. Memory and processing power are restricted on these computers.

Big data Visualization analytics

Big data Visualization analytics involves a computer-analytics methodology that combines data mining, statistics, and visualization. In addition to automated processing, visual analytics systems include human input to allow for user control and judgment during data analysis, resulting in useful insight for decision-making [7,8]. Multiple visual analytics analysis projects have been undertaken throughout the years. The majority of them work with a standard visual analytics pipeline, which portrays the visual analytics cycle. The visual analytics phase begins by executing data processing sub-tasks including sorting and filtering, which turn the data collection into expressions that can be explored further.

The pipeline adopts either a visual discovery approach or an automated analysis method, based on the particular usage case. For automated processing, data mining techniques are used to help characterize the data. Analysts and decision-makers use the visual interface to interpret and evaluate the results [8].

Data, Visualization, Models, and Knowledge are the four key principles of the Visual Analytics Pipeline architecture. The data component is responsible for unprocessed and heterogeneous data storage as well as pre-processing. Because data collection is carried out by sensors in real-time, the raw data sets are typically fragmented, distorted or inaccurate and are not specifically included in the visualization framework or the model's module. Any pre-processing data needs to be added to the initial data sets to remove these problems. Data pre-processing, based on raw data content, is a scalable operation., data integration, Pre-processing strategies like data parsing, data cleaning (removing clutter, bugs, and invalid data), the transformation of data (normalization), and data reduction are used in this module. The Models module converts data into facts. This module involves conversion techniques such as selection and creation of functionality, model development, selection, and validation.

D. IoT Data Visualization Tools

Visualization applications enhance the decision-making process because they have powerful data analytics that can visualize massive data obtained from multiple IoT devices. IoT data visualizing systems have a custom dashboard interface that helps the user to analyze raw metrics available and obtain information on the functionality of the models, considering the collection of observations made from many geographically dispersed IoT sensors as well as several AI models added to the results [9]. The key purpose of these programs is to increase the operator's confidence in the models. A modular visualization framework must retain some key features, including the ability to update features in real-time, interactive elements, openness, and clarity. Since the IoT metrics are extremely complex and new measurements are taken in full detail, visualizations need to refresh when new measures are accessible in real-time [9]. The dashboard can have an immersive user experience that allows operators to access and communicate with data. The dashboard can also include ways to view and simulate the implemented AI models, to improve the clarity and interpretability of the models.

Most of the visualization frameworks are developed with SOA, including 4 major offerings: a data collection, a data display a dynamic dashboard service, which establishes a connection that helps organize and displays the different combination of text, machine value, or the As a service, large-scale data systems, the big-scale data framework, and big-data processing tools [9]. In this segment, the most commonly employed IoT data visualization applications are listed in many major industries. Such requirements were related to the following: open-source software, the potential to connect with common data sources (for instance Google Analytics, MapR Hadoop Hive, Cloudera Hadoop, Salesforce, etc.," graphical visualization, consumer category (desktop, online or smartphone app).

Tableau is a simple and scalable data visualization platform that facilitates user engagement. Its user interface offers a wide array of sets and personalized visualizations using a large selection of intuitive charts. Thorough analyzes may be performed with R-scripting. It supports most data types and links to different servers, including Amazon Aurora, Cloudera Hadoop, and Salesforce. The web service of Tableau is open to the public but requires restricted storage. Server and laptop models with commercial licenses are eligible.

ThingsBoard is an open-source IoT framework for system control, data storage, encoding, and viewing modules. The framework enables the development of customized IoT dashboards with widgets to view visual data gathered from multiple devices. It provides a variety of features, like line and bar chart components, for chronological and real-time data displays [11]. It also includes map widgets to track objects on online charts. Its complicated stack technology (Java, Python, C++, JavaScript) offers error-free output and data analysis in real-time. It follows common Internet of Things protocols for system networking (e.g. MQTT, CoAP, and HTTP) [10]. It can be incorporated with a customized feature with Node-Red, a flow-based IoT programming platform [11].

Plotly is a shared data analysis service focused on the online cloud. It is created with the Python and Django frameworks. It offers numerous IoT visualization and analytics data management facilities and modules. It permits the development of online dashboards with a broad variety of charts such as mathematical, analytical, 3D charts, multi-axis charts, etc. It offers in-depth analysis APIs focused on Python, R, MATLAB, and Julia [11]. Furthermore, graphical libraries including ggplot2, matplotlib, and MATLAB map conversion technology improve visualization. The Web Plot Digitizer (WPDinternal)'s application will automatically record data from static photographs. It is open to the public with restricted map functionality and storage and the whole range of chart characteristics is accessible under a specialist membership license [11].

IBM Watson IoT Platform is a software architecture that supports many programming languages, services, and advanced DevOps in the deployment and management of cloud applications. It contains a series of optimized web apps, thus enabling device integration by REST APIs for third-party applications. Static and dynamic data are visualized simply by constructing independent diagrams, tables, and graphs. It offers links to system requirements and control of warnings. For IoT computer connectivity, APIs, and online resources, Node-RED can be used. For more data processing, sensor data deposited in Cloudant NoSQL DB can be analyzed. Power BI is a popular cloud-based market analytics service. It offers a diverse collection of immersive views and comprehensive analytical reports for major companies. It is programmed to trace and visualize different data obtained by the sensor [12]. The framework integrates for cloud-based visualization and cognitive services from Azure. It includes three basic components: Power BI Desktop, Report Generator, Service (SaaS), document publishing, and Apps [12]. Numerous forms of source integrations are provided as well as rich data visualizations. Data may be reviewed utilizing the natural language query function and other tools. Data processing is carried out in both real-time and static historical data streaming. Power BI offers subcomponents for IoT incorporation.

Grafana is an upcoming data visualization platform for open-source sources. It emphasizes the analysis and tracking of metric data over time. As a result, it works for a variety of time-series data storage backends [13]. It offers real-time feeds with almost real-time data collection reports. The versatile and immersive visual dashboard encompasses numerous features, facilitates very personalized questions, and includes several graphics to analyze and display data, including heat maps, histograms, geo maps, charts, graphs. Grafana integrates seamlessly different data sources such as Elasticsearch, MySQL, InfluxDB, PostgreSQL, Prometheus, Graphite, OpenTSDB, and KairosDB. It offers cloud storage that allows full network access. It also offers warning messages in the likelihood of an adverse event [13].

Kibana Platform is an open-source data visualization tool specialized in the broad collections data processing. It is part of the ELK Stack, a collection of three instruments: ElasticSearch, LogStash, and Kibana. While Elasticsearch and Logstash could work separately, they were created as an integrated system. Logstash is in charge of gathering data from multiple remote outlets, while Elasticsearch offers storage resources to Kibana, which is in charge of data visualization. Beats, which is the fourth product introduced to the stack, transmits messages from the network edge to Logstas as a lightweight supply [14].

Additionally, a Cyber-Physical System (CPS) is a group of hardware devices linked to simulated cyberspace via a transmission line. Each hardware device is connected to a cyber model which stores all knowledge and information [15]. This cyber paradigm is referred to as the Digital Twin. It facilitates the transmitting of data from the physical to the cyber component. Nevertheless, the spatial time relationship between each digital twin is far more beneficial than the real digital twin in a few CPS where each physical entity has a digital twin equivalent. Digital twins can be produced using 3d animation using AR/VR/MR and even hologram technology. Digital twins combine advanced techniques like haptics, humanoid robotics, soft robotics, humanoid robotics,5G, cloud computing, and IC [16].

IV. IT'S FUTURE IN USA

The future of U.S. Big Data Visualization in the IoTs significantly increases productivity and boosts quality by providing infographics that businesses can use to gain useful insights. As competition for international markets becomes challenging especially from countries like China, U.S companies are already employing big data analytics to predict the future. Nothing beats being able to share perspectives using interactive graphics in real-time. The U.S healthcare system is already benefiting from the technology [17,18]. The number of smart health tracking systems has grown significantly in recent years. These technologies can create vast quantities of data. visualizations, Therefore, using big data visualizations to interpret data from fetal monitoring, temperature monitors, electrocardiograms, temperature monitors, or level of blood glucose monitors support healthcare practitioners in rapidly evaluating patients' physical conditions. Furthermore, big data visualizations help healthcare workers to identify serious illnesses early on, possibly saving lives.

The Internet of Things (IoT) is projected to play a significant role as an evolving technology in the United States, with added advantages in the areas of retail and logistics. RFID is used in logistics to track containers, pallet trucks, and barrels. Furthermore, major advances in IoT technology support retailers by offering a range of advantages [18]. The major US players in this industry, such as IBM, Google, and Microsoft, are pursuing a merger and acquisition strategy to extend their global presence and increase sales in the coming years. Also, they are growing in cutting-edge technology innovations such as artificial intelligence, natural language processing, and others to retain a competitive advantage in the industry.

V. BENEFITS TO U.S ECONOMY

The application of big data visualization for IoT devices will transform the U.S economy tremendously. This technology is being implemented by companies in a wide range of markets, highlighting its long-term potential. Even at this early stage of IoT deployment, the advances in performance, automation, predictive analysis, system integration, and making predictions have resulted continuously point toward an economy that is more effective, sustainable, secure, and profitable. At the end of the decade, it may be worth up to $15 trillion in economic valuation. To put this in context, the United States' overall total gross domestic product is $17.9 trillion. Industries (and also private entities) that have unexplored the big data visualizations in IoT will become increasingly rare in the near future. IBM is at the frontline of assisting clients in introducing emerging innovations that not only enhance operating performance but also recast consumer interactions, revolutionize market strategies, and enable firms to reimagine how their markets work [18]. Daimler has evolved to be one of the world's largest car manufacturers, manufacturing well-known brands such as Mercedes, Smart, Maybach, and Freightliner ever since its establishment in 1924. Daimler, dissatisfied with only using IoT technology to transform internal activities, enlisted IBM's support to deliver car2go, an on-demand fleet of eco-friendly Smart cars that consumers can book through a smartphone app. Car2go is a radical reinvention of the automaker's position in the automotive sector, and none of it will be feasible without IoT big data visualizations.

VI. CONCLUSION

The advent of IoT services dramatically accelerated the growth pace at which vast and complicated data sets were created. Integrating human judgment into the methods for data processing allows visual analysis to uncover information and to obtain useful perspectives from this data collection. In this respect, each piece of IoT data is critical for retrieving information and identifying valuable trends. When data is interpreted visually, human perceptual and cognitive functions reliably distinguish trends. When it comes to processing voluminous and interpreting IoT data without losing efficiency and reaction time, data visualization approaches are very important.

References

[1] G. Andrienko, N. Andrienko, P. Bak, D. Keim, S. Wrobel: Visual Analytics of Movement. Springer 2013.

[2] A.F. Batista, P.L. Correa and G. Palanisamy, "Visual Analytics Improving Data Understandability in IoT Projects: An Overview of the US DOE ARM Program Data Science Tools", *Mobile Ad Hoc and Sensor Systems (MASS) 2016 IEEE 13th International Conference on*, pp. 349-354, 2016, October.

[3] H. Wang, Z. Xu and W. Pedrycz, "An overview on the roles of fuzzy set techniques in big data processing: Trends challenges and opportunities", Knowledge-Based Systems, vol. 118, pp. 10-12, 2016.

[4] D. Keim, G. Andrienko, J.D. Fekete, C. Görg, J. Kohlhammer and G. Melançon, "Visual analytics: Definition process and challenges" in Information visualization, Berlin, Heidelberg:Springer, pp. 154-175, 2008.

[5] N. Bikakis, T. Sellis: Exploration and Visualization in the Web of Big Linked Data: A Survey of the State of the Art, LWDM Workshop 2016.

[6] E.T. Brown, A. Ottley, et al.: Finding Waldo: Learning About Users from their Interaction TVCG 20(12), 2014

[7] C. D. Correa, Y.-H. Chan, K.-L. Ma: A framework for uncertainty-aware visual analytics. VAST 2009

[8] L. Wang, G. Wang and C.A. Alexander, "Big data and visualization: methods challenges and technology progress", *Digital Technologies*, vol. 1, no. 1, pp. 33-38, 2015.

[9] T. Fujiwara, O.-H. Kwon, K.-L. Ma: Supporting Analysis of Dimensionality Reduction Results with Contrastive Learning. TVCG 26(1), 2020

[10] P. Godfrey, J. Gryz, P. Lasek: Interactive Visualization of Large Data Sets. TKDE 28(8), 2016

[11] D. Keim, J. Kohlhammer (eds.): Mastering the Information Age: Solving Problems with Visual Analytics. Eurographics 2010

[12] A. Lior, J. Allen, O. Barykin, V. Borkar, B.Chopra, et al.: SCUBA: diving into Data at Facebook. PVLDB 6(11), 2013

[13] Y. Lou, R. Caruana, J. Gehrke, G. Hooker: Accurate Intelligible Models with Pairwise Interactions. KDD 2013

[14] B. Mutlu, E.E. Veas, C. Trattner VizRec: Recommending Personalized Visualizations. TiiS 6(4), 2016

[15] M. Savva, N. Kong, A. Chhajta, L. Fei-Fei, M. Agrawala, J. Heer: ReVision: Automated Classification, Analysis and Redesign of Chart Images. UIST 2011

[16] Y. Wang, K.-L. Ma: Revealing the fog-of-war: A visualizationdirected, uncertainty-aware approach for exploring high-dimensional data. IEEE BigData 2015

[17] Y. Wu, G.-X. Yuan, K.-L. Ma: Visualizing Flow of Uncertainty through Analytical Processes. TVCG 18(12), 2012

[18] T. Zuk, S. Carpendale: Theoretical Analysis of Uncertainty Visualizations. Visualization and Data Analysis, 2006