Journal of Harbin Engineering University
ISSN: 1006-7043

Vol 45 No. 11
November 2024

Internet of Everything Advancement Study in Data Science and Knowledge Analytic Streams

[1]Mr. Achinta Kumar Palit, [2]Prof(Dr.) Bibhuti Bhusan Behera (HOD)
[1]Assistant Professor in CSE,
Gandhi Institute of Excellent Technocrats, Bhubaneswar
[2]Assosiate Professor in CSE, Gandhi Institute of Excellent Technocrats, BhuBaneswar
[*]Sushree Subhasmita Mohanty Reseach Scholar, BPUT
mohantysubhasmita1129 @ gmail.com

Abstract: Internet of Everything (IoE) generates a massive amount of diverse data. To manage this, data is stored using a special method called a "column-oriented relational framework." In IoE databases, there are lots of rows (data points) but fewer columns (data types). A column-oriented approach helps improve performance and manage this large amount of data more efficiently. The way data is stored in these systems can sometimes lead to problems like data inconsistency and integrity issues, especially when dealing with different types of data in the same database. Using a column-oriented framework helps address these issues and makes it easier to store and access data quickly. Analytics is crucial in data science. It helps in understanding and using the data effectively. Since technology and business needs keep changing, analytics is an ongoing process. Two data scientists might solve the same problem in different ways, showing how flexible and creative analytics can be. The main goals are to create and study new data models, frameworks, and algorithms for handling IoE data. These advancements aim to improve how we manage IoE databases and discover valuable insights from the data. The research focuses on improving the way IoE data is stored and managed using advanced techniques, addressing storage challenges, and enhancing data analysis to adapt to ever-changing needs.

Keywords:—column oriented database, Internet of Everything (IoE) database, knowledge analytic, data depository, data science

Introduction

The IoE database management system is a platform for handling large amounts of structured, unstructured, and semi-structured data. In many IoE applications, a lot of data is stored for later analysis to find useful insights. For example, sensors in an aircraft produce a huge amount of data, around 20 terabytes per hour, which is saved in a black box. This data is then accessed from the black box as needed for further analysis. While digital storage technology has greatly improved to handle large data, the speed of accessing and analyzing this data still has limitations. So, managing time-sensitive data access and performing predictive analysis on mostly unstructured IoE data remain challenging problems.

The Internet of Everything (IoE) has a major impact on Big Data. A key point in the evolution of IoE data science is that each IoE object has a unique identifier and is connected to others. With potentially trillions of these connections generating massive volumes of data (IoE big data), current data science and analytics methods will face significant challenges. The IoE network links people, processes, places, and things to the internet for communication across the globe. IoE objects include both physical and logical elements. Physical things are people, locations, and devices, while logical things are processes, frameworks, applications, software, and programs. All this data structured, semi-structured, and unstructured forms a comprehensive IoE database [1].

In an IoE database, different types of data are categorized as follows:

- Structured Data: ERP and CRM data.
- Semi-Structured Data: XML data.
- Unstructured Data: Email documents, social media content, PDFs, Word documents, and rich text documents.

Studies show that about 80% of data in an IoE database is unstructured and lacks predefined data models. This unstructured data can include text, graphics, videos, and symbols. Additionally, spatio-temporal databases, which contain facts or events with time stamps, are part of the IoE database. The rapid growth of IoE big data applications today leads to several problems, including issues with data volume and speed.

Analyzing and drawing insights (knowledge) from large IoE databases in real-time is becoming increasingly challenging due to the rapid growth in data volume and variety associated with many IoE applications. Real-time IoE knowledge analytics face several problems, including:

- Managing diverse types of knowledge
- Converting various types of data into useful knowledge
- Turning that knowledge into actionable steps
- Using those actions to make informed decisions
- Adjusting decisions to effectively coordinate IoE applications [2]

Researchers have explored combining statistical and computational learning methods to address problems in data science and knowledge analytics. These methods are used to handle various tasks in IoE big data, including data transformation and analysis, data mining, knowledge discovery, semantic exploration, and structural analysis. Machine learning techniques are applied in many areas of knowledge discovery and semantic analytics to enhance application intelligence. However, in most IoE big data applications, a large volume of data is stored that is often redundant and not ideal for analysis, modelling, information transformation, knowledge creation, or decision-making. A survey by Par Stream reveals that 94% of organizations face challenges with IoE big data management and analytics. Despite these challenges, 70% of organizations believe that IoE big data analytics can lead to better and more meaningful decisions[3], [18].

The main goals of the research are to design and explore data models, frameworks, architectures, and algorithms focused on network-centric data, especially IoE data, to support data science and knowledge analytics tasks for intellectual domain applications.

The objective is to create knowledge analytics frameworks and algorithms that can build and organize knowledge and insights. This process aims to turn real world applications into intellectual domain applications.

Data science and knowledge analytics are becoming increasingly popular for various intellectual domain applications. The goal is to derive valuable insights from large-scale network-centric data, such as IoE data, to generate intelligence for these applications. Currently, network-centric data in intellectual domain applications are often unstructured and ambiguous, posing challenges for extracting useful knowledge. Surveys show that for IoE big data applications, the time to gain insights is slow, the quality of insights is poor, and the cost is high. In contrast, intellectual domain applications need low-cost, high-quality, and real-time frameworks and algorithms to efficiently transform their data into valuable cognitive insights. These insights are used to enhance the value of intellectual domain applications.

Related Studies

Here is a summary of the evolution of data science and analytics (DSA) through its key developments and milestones:

1960: Peter Naur uses "data science" as a term similar to computer science.

1974: Naur refers to data science as methods for data processing across various applications [4].

1977: Turkey suggests that data science involves exploratory data analysis [5], [23].

1989-1996: New terms and tasks are added to data science, including data classification, data mining, and knowledge discovery [6].

1997-2001: Statistical computing becomes part of data science.

2005: Thomas H. Davenport and others introduce the use of analytics and data-based decision-making [7], [26].

2010: Hilary Mason and Chris Wiggins introduce the term "machine learning" in the context of data science.

2011: Harlan Harris discusses various data science techniques, such as statistics, machine learning, data interpretation, classification, and visualization [8], [26].

2012 to Present: Data science integrates with emerging technologies, including IoE, big data, cloud computing, deep learning, extreme learning machines (ELM), and other advancements [23], [24], [25].

Data science is increasingly focused on handling both big data processing and knowledge analytics, which are key challenges due to the growing scale and diversity of data from IoE-driven intellectual domain applications. IoE data architecture typically uses NoSQL databases and Hadoop-like platforms for batch processing of large datasets, which can be time-consuming. However, real-time or semi-real-time data processing, management, and knowledge analytics are even more complex tasks. Current business intelligence platforms require timely insights to turn business data into valuable, actionable knowledge to drive revenue and reduce potential risks. NoSQL databases are not designed to handle the knowledge analytics tasks needed for IoE-driven applications, which are essential for deriving meaningful insights and making informed decisions.

Column Oriented Relational Framework

The IoE database is a spatio-temporal database that includes event data with timestamps and geographical locations. For a given time and location, the physical data value is accurate, but this accuracy may not apply universally to all locations and timestamps. For example, at location L001 and timestamp T001, the environmental temperature might be 40 degrees Celsius. In a large-scale IoE database, there are many rows and relatively few columns. Using a column-oriented relational framework can significantly enhance the performance of the IoE database, particularly for data storage and access management. Consider an IoE relational database 'R' with the following physical schema.

In the relation 'R' (Table-1), the following functional dependencies are defined:

FD1: IoE Object ID, Timestamp → Temperature

FD2: IoE Object ID, Timestamp → Humidity

This means that, for a given IoE Object ID and Time stamp, the Temperature and Humidity values are uniquely determined. In the given functional dependencies, IoE Object ID and Timestamp together serve as a composite primary key for relation 'R'. This composite key uniquely identifies each record in the database.Additionally, within the geographical coordinate system, the Location ID can be uniquely determined by the IoE Object ID. This means that for each IoE Object ID, there is a specific Location ID associated with it.

Table1. Sample IoE relational data in R

IoE object ID	*Location ID*	*Timestamp*	*Temp(Celsius)*	*Humidity (%)*
I001	L001	T001	29	48
I002	L002	T001	30	49
I003	L001	T002	34	53
I004	L002	T002	39	57
I005	L003	T001	31	68

Now, we can express, R‟ into a column oriented relational framework as follows.

[{I001,T001:R001};{I002,T001:R002};{I003,T002: R003}…]

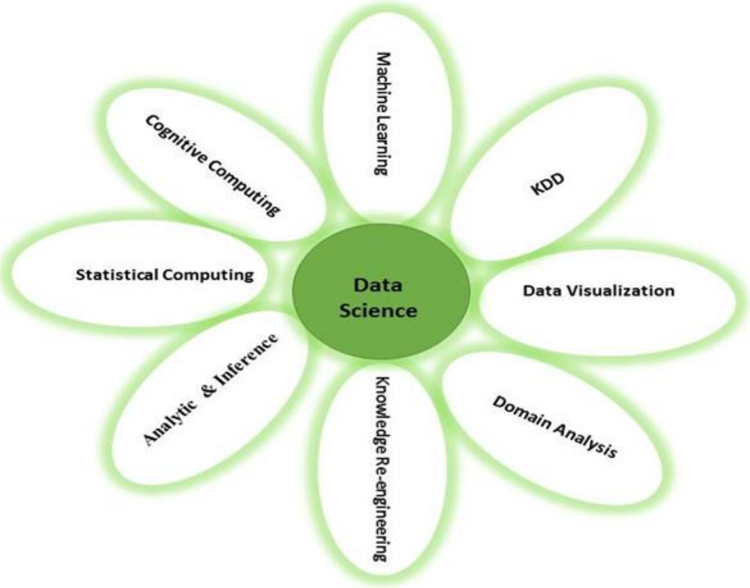
[{29:R001};{30:R002};{34:R003}…]

[{48:R001};{49:R002};{53:R003}…]

Where R001, R002… are the row identifiers and uniqueness is maintained for each row of relation „R‟.

The multi-object IoE device is equipped with various sensors that generate diverse types of physical data in different formats and structures. This variety leads to challenges in data extraction and classification, and it also complicates predictive knowledge analytics.

Context of Data Science and Analytics



Science involves gaining knowledge through systematic study, so data science similarly focuses on analyzing data. This includes studying the organization, properties, and analysis of data, as well as understanding how data informs our conclusions and confidence in those conclusions [10]. The promise of data science is that by recording and understanding data from a system, we can use this knowledge and these insights to improve the system. Data science involves extracting knowledge from both structured and unstructured data. It builds on the fields of data mining and predictive analytics, which are also referred to as knowledge discovery and data mining [11].

In data science, various tasks are involved that can be explored and integrated to develop applications for different intellectual domains connected to the physical world. Figure 1 illustrates the structure of data science. Key tasks in data science include:

Statistical Computing and Visualizations: These tasks cover data manipulation and cleaning, importing and exporting data, managing missing values, handling data frames, functions, lists, and matrices, writing functions, and using packages. Efficient programming practices and methods for summarizing and visualizing data are crucial.

Cognitive Computing: This area of data science focuses on creating new computational methods to tackle complex problems through self-learning processes. Cognitive computing uses a range of technologies, such as mathematics, statistics, data science, and computational science, to generate intelligent insights and solutions for intellectual domain applications.

IoE objects can be either logical or physical entities related to real-world objects. Logical things include processes, frameworks, software, and applications, while physical things include people, places, and devices, as shown in Figure 2.

To construct an IoE object, three main requirements are needed:

- Physical IP System: Provides a unique identity for each object.
- Radio Transceivers: Facilitate communication using protocol stacks.
- Sensing Unit: Collects data from the physical environment.

The goal is to integrate these components seamlessly into real-world entities connected to IoE applications. Such connectivity results in a large and diverse collection of data, with inconsistencies among different database frameworks. These inconsistencies—such as differences in naming conventions, scale, structure, and levels of abstraction—pose significant challenges for exploring and analyzing IoE data. As IoE-based applications rapidly grow, networks of billions of IoE objects generate vast amounts of structured, semi-structured, and unstructured big data in real time. This proliferation of data creates complex environments for IoE data science, where managing and analyzing this data becomes increasingly challenging.

Based on risk quantification, real-time data science applications can be classified into three categories:

- **Business-Critical Applications:** These include IoE applications used in business intelligence monitoring. They are important for business operations but generally involve lower risk compared to the other categories.
- **Mission-Critical Applications:** These applications are vital for specific missions or functions, such as habitat monitoring, smart city monitoring, and smart home monitoring. They are crucial for maintaining the effectiveness of systems and services but carry moderate risk.
- **Safety-Critical Applications:** These involve IoE applications in areas like industrial automation, healthcare automation, and elderly activity supervision. They carry the highest degree of risk due to their impact on safety and well-being.

Among these categories, safety-critical IoE applications are considered the highest risk due to their potential consequences on human safety and health [46].

In the convergence of IoE and Data Science and Analytics (DSA), several emerging technologies are being progressively integrated. In our work, we utilize data from the Internet of Everything (IoE) to analyze case-based problem scenarios and model applications. When examining emerging technologies, we find that many are currently in the "innovation trigger" phase, where initial breakthroughs are made, but widespread adoption is still developing. Additionally, several convergence technologies, including Data Science and Analytics, cloud computing, the Internet of Things (IoT), IoE, computational learning, and range-based natural language querying, are at the "peak of inflated expectations." At this stage, these technologies are receiving a lot of hype and interest, but practical implementation and benefits may still be emerging[35].

Conclusion and Future Scope

This discussion has explored the evolving operations of data science and knowledge analytics. I have covered various mechanisms for analyzing and exploring tasks such as data transformation and analysis, data mining, knowledge discovery, semantic knowledge exploration, and structural analysis. Machine learning techniques are applied in many areas of knowledge discovery and semantic analytics to enhance application intelligence. The main goal is to derive insights from large-scale IoE data to generate valuable intelligence for applications. Future work will focus on developing advanced data science and knowledge analytics frameworks and applications. These advancements aim to unlock significant value from diverse data sources, regardless of their scale. Managing and analyzing large-scale, disparate IoE databases remains a significant challenge, highlighting the need for continued innovation in data science and analytics.

References

[1] Mishra,Nilamadhab,Chung-ChihLin,andHsien-TsungChang. "A Cognitive Oriented Framework for IoT Big-data Management Prospective."Communication Problem-Solving (ICCP), 2014IEEE International Conference on.IEEE, 2014.

[2] S.J. Nasti, M. Asgar, M.A. Butt , "Analysis of CustomerBehaviour using Modern Data Mining Techniques", International Journal of Computer Sciences and Engineering, Vol.5, Issue.12, pp.64-66,2017.

[3] Zarko, IvanaPodnar, et al. "IoT data management methods and optimisation algorithms for mobile publish/subscribe services in cloud Environment. “Networks and Communications (EuCNC), 2014 European Conference on. IEEE, 2014.

[4] Gubbi, Jayavardhana, et al. "Internet of Things (IoT): A vision, architectural elements, and future directions." Future Generation Computer Systems 29.7 (2013): 1645-1660.

[5] Anantharam, Pramod, PayamBarnaghi, and AmitSheth. "Data Processing and Semantics for Advanced Internet of Things (IoT) Applications: modeling, annotation,

integration, and perception." Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics. ACM, 2013.

[6] Mishra, Nilamadhab, Hsien-Tsung Chang, and Chung-Chih Lin. "An IoT Knowledge Re-engineering Framework for Semantic Knowledge Analytics for BI-services." Mathematical problems in engineering, vol. 2015, Article id-759425, 12 pages (2015).

[7] Dagnino, Aldo, and David Cox. "Industrial Analytics to Discover Knowledge from Instrumented Networked Machines."Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering (SEKE‟14), Vancouver, Canada.2014.

[8] Chang, H. T., Mishra, N., & Lin, C. C., IoT big-data centred knowledge granule analytic and cluster framework for BI applications: a case base analysis. PloS one, 10(11), 2015.

[9] Mishra, N., Chang, H. T., & Lin, C. C. Sensor data distributionand knowledge inference framework for a cognitive-based distributed storage sink environment. International Journal of Sensor Networks, 26(1), 26-42,2018.

[10] Mishra N,."In-network Distributed Analytics on Data-centric IoT Network for BI-service Applications", International Journal of Scientific Research in Computer Science, Engineering and InformationTechnology(IJSRCSEIT), ISSN: 2456-3307, Volume 2, Issue 5, pp.547-552, September-October.2017.

[11] Patnaik, B. C., & Mishra, N. “A Review on Enhancing the Journaling File System”, Imperial Journal of Interdisciplinary Research, 2, no. 11 (2016)

[12] Chang, H. T., Yu-Wen Li., & Mishra, N. “mCAF: A Multi- dimensional Clustering Algorithm for Friends of Social Network Services”, Springer Plus, 2016.

[13] Chang, H. T., Liu, S. W., & Mishra, N. “A tracking and summarization system for online Chinese news topics”, Aslib Journal of Information Management, 67(6), 687-699,2015.

[14] Mishra, N., Lin, C. C., & Chang, H. T. “A Cognitive Adopted Framework for IoT Big-Data Management and Knowledge Discovery Prospective”, International Journal of DistributedSensor Networks, 2015.

[15] Mishra, N., Lin, C. C., & Chang, H. T.“Cognitive inference device for activity supervision in the elderly”, The ScientificWorld Journal, 2014.

[16] Mishra, N., Chang, H. T., & Lin, C. C.“Data-centric Knowledge Discovery Strategy for a Safety-critical Sensor Application”, International Journal of Antennas and Propagation, Article ID 172186, 11 pages, 2014. doi:10.1155/2014/172186.

[17] Cheng, Hsu-Chen, and Wen-Wei Liao. "Establishing a lifelong learning environment using IOT and learning analytics."Advanced Communication Technology (ICACT), 2012 14th International Conference on. IEEE, 2012.

[18]http://sites.parstream.com/parstream-iot-survey-whitepaper(survey march 2015)

[19] Chawla, S., Hartline, J., &Nekipelov, D. , Mechanism design for data science. In Proceedings of the fifteenth ACM conference on Economics and computation (pp. 711-712). ACM, 2014-june.

[20] https://en.wikipedia.org/wiki/Data_science

[21] Provost, F., & Fawcett, T., Data Science for Business:What you need to know about data mining and data-analytic thinking. "O'Reilly Media, Inc." 2013.

[22]http://www.businessdictionary.com/definition/analytics.htm.

[23]http://www.forbes.com/sites/gilpress/2013/05/28/a-very-short- history-of-data-science/

[24] Tukey,J.W.,We need both exploratory and confirmatory.The American Statistician, 34(1), 23-25, 1980.

[25] Hayashi, C., Yajima, K., Bock, H. H., Ohsumi, N., Tanaka, Y., & Baba, Y. (Eds.). (2013). Data Science, Classification, and Related Methods: Proceedings of the Fifth Conference of the International Federation of Classification Societies (IFCS-96), Kobe, Japan, March 27–30, 1996. Springer Science & Business Media.

[26] Davenport, T. H., Harris, J. G., & Cantrell, S., Enterprise systems and ongoing process

change. Business Process Management Journal, 10(1), 16-26, 2004.

[27] Shron, M. (2014). Thinking with Data: How to Turn Information Into Insights. " O'Reilly Media, Inc.".

[28]https://en.wikipedia.org/wiki/Internet_of_Things.

[29] http://www.scaledb.com/internet-things-database.php

[30] Cheng, Hsu-Chen, and Wen-Wei Liao. "Establishing a lifelong learning environment using IOT and learning analytics."Advanced Communication Technology (ICACT), 2012 14th International Conference on. IEEE, 2012.

[31] Zarko, IvanaPodnar, et al. "IoT data management methods and optimization algorithms for mobile publish/subscribe services in cloud environments. "Networks and Communications (EuCNC), 2014 European Conference on. IEEE, 2014.

[32] Gubbi, Jayavardhana, et al. "Internet of Things (IoT): A vision, architectural elements, and future directions." Future Generation Computer Systems 29.7 (2013): 1645-1660.

[33] Anantharam, Pramod, PayamBarnaghi, and AmitSheth. "Data Processing and Semantics for Advanced Internet of Things (IoT) Applications: modeling, annotation, integration, and perception." Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics. ACM, 2013.

[34] Dagnino, Aldo, and David Cox. "Industrial Analytics to Discover Knowledge from Instrumented Networked Machines."Proceedings of the 26th International Conference on Software Engineering and Knowledge Engineering (SEKE‟14), Vancouver, Canada.2014.

[35]Dou,Dejing,HaoWang,andHaishanLiu."Semanticdatamining: A survey of ontology-based approaches." Semantic Computing (ICSC), 2015 IEEE International Conference on. IEEE, 2015.

[36] Philip Chen, C. L., & Zhang, C. Y., Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Information Sciences 2014, 275, 314-347.

[37] Mehta, A., Big Data: Powering the Next Industrial Revolution. Tableau Software White Paper 2011.

[38] Comas, D. S., Pastore, J. I., Bouchet, A., Ballarin, V. L., &Meschino,G.J.,Type-2FuzzyLogicinDecisionSupport

[41] Sheta, A. F., Braik, M., Öznergiz, E., Ayesh, A., &Masud, M., Design and Automation for Manufacturing Processes: An Intelligent Business Modeling Using Adaptive Neuro-Fuzzy Inference Systems. In Business Intelligence and Performance Management 2013 (pp. 191-208). Springer London.

[42]Makridou,G.,Atsalakis,G.S.,Zopounidis,C.,&Andriosopoulos, K., Gold price forecasting with a neuro–fuzzy–based inference system. International Journal of Financial Engineering and Risk Management, 2013, 1(1), 35-54.

[43] Lemos, A., Caminhas, W., &Gomide, F., Multivariable gaussian evolving fuzzy modeling system. IEEE Transactions on Fuzzy Systems, 2011, 19(1), 91-104.

[44] F. Hair Jr, J., Sarstedt, M., Hopkins, L., & G. Kuppelwieser, V. , Partial least squares structural equation modeling (PLS-SEM) An emerging tool in business research. European Business Review, 2014, 26(2), 106-121.

[45] Chen, Y., & Chiu, Y., Enhancing business intelligence for supply chain operations through effective classification of supplier management. Uncertain Supply Chain Management, 2014, 2(4), 229-236.

[46] Al-Shayea, Q. K., & Al-Shayea, T. K., Customer Behavior on RFMT Model Using Neural Networks. In Proceedings of the World Congress on Engineering 2014 (Vol. 1).

[47] Al-Shayea, Q. K., & El-Refae, G. A., Evaluation of banks insolvency using artificial neural networks. In Proceedings of the 11th WSEAS international conference on Artificial Intelligence, Knowledge Engineering and Data Bases (AIKED'12), Cambridge, United Kingdom, 2012 February (pp. 22-24).

[48] Wu, D. (2012). On the fundamental differences between interval type-2 and type-1 fuzzy logic controllers. Fuzzy Systems, IEEE Transactions on, 20(5), 832-848.

[49] Adderley, R., Seidler, P., Badii, A., Tiemann, M., Neri, F., &Raffaelli, M. (2014, July).

Semantic Mining and Analysis of Heterogeneous Data for Novel Intelligence Insights. In IMMM 2014, The Fourth International Conference on Advances in Information Mining and Management (pp. 36-40).

[50]BhanudasSureshPanchabhai,AnandJayantilalMaheshwari,Sunil DhonduMone, "The Road Map of Cloud Computing to Internet of Things",Notice:Undefinedindex:jour_sub_namein/home/isrosa2l/public_html/journal/IJSRCSE/special_issue.phpon line 283, Vol.06, Issue.01, pp.37-42, 2018.