CONTENTS

A Review of Data Mining Techniques for Enhancing Energy Efficiency in IoT Systems

[1] *Dr. S. Sharmila,*
[1]Assistant Professor, UG Department of Computer Science (SF),
Nallamuthu Gounder Mahalingam College, Pollachi, Tamilnadu, India.

[2] *Dr.A.Kanagaraj,*
[2]Associate Professor, Dept. of Computer Science,
Kristu Jayanti Deemed to be University, Bengaluru, India,

Abstract

The rapid proliferation of Internet of Things (IoT) technologies has led to an unprecedented surge in data generation and energy consumption, raising concerns about sustainability and operational efficiency. This review explores the role of data mining techniques in enhancing energy efficiency across various IoT applications. By analyzing and classifying key data mining methods such as clustering, classification, association rule mining, and anomaly detection this paper highlights how intelligent data processing can optimize energy usage in IoT networks, reduce resource wastage, and prolong device lifespan. The review further discusses real-world case studies, emerging trends, and challenges in integrating data mining with energy-aware IoT systems. It concludes by identifying potential research directions aimed at building smarter, greener, and more sustainable IoT infrastructures.

Keywords: Internet of Thongs (IoT), Energy, Clustering, Classification, Efficiency.

1. INTRODUCTION

The Internet of Things (IoT) refers to the type of the network which connect anything i.e. physical objects-devices, buildings, vehicles and other items embedded with software, sensors and network connectivity based on stipulated protocols that enables these objects to collect and exchange data. In our daily lives, we have become more reliant on IoT with our wearable tech, appliances, our cars, how we receive health care. Due to Seamless integration of classical networks with IoT, it enables a great vision that all things can be easily monitored and controlled which results in to voluminous data. So, in order to make IoT more smarter, lots of data analysis is needed for which one of the most solution is data mining [2]. Much research in recent years has focused on data mining in Internet of Things (IoT) which connects physical objects, person to person, person to machine or machine to machine via internet and manages information.

Data mining process refers to the process of semi automatically analyzing large databases for pattern mining which are innovative, legitimate, useful and understandable which is also known as Knowledge Discovery in Databases (KDD) [3]. Data mining or KDD process includes problem formulation, data collection, data cleaning i.e. pre-processing, transformation, choosing mining task/method and result evaluation/visualization. Knowledge discovery is an iterative process. Data mining overlaps with other fields like statistics, machine learning, artificial intelligence, databases but mainly it focuses on automation of handling large heterogeneous data, algorithm and scalability of number of features and instances. The objective of any data mining process is to build an efficient predictive or descriptive model of a large amount of data that not only best fits or explains it, but is also able to generalize to new data [14].

At first in 1999 The Massachusetts Institute of Technology (MIT) put forwarded the concept of IoT. IoT is described as data and things round the clock connected through the Internet. The IoT also views everything as the same, as things. These things includes Smartphones, users, processing units tablets, Bluetooth, ZigBee, data centers (DCs), the Infrared Data Association (IrDA), cellular networks, ultra-wideband (UWB), near field communication (NFC) DCs, Wi-Fi networks, RFID, chips and sensors, vehicles, wristwatches, household equipments means IoT is mixture of factual things and virtual things which are connected anytime and anywhere.The data generated or collected by these devices is huge in volume. This data will be in vast amounts for a system, and immense for a larger systems. To conserve and produce significant business information out of this data, to provide different services to improvement of the business growth and system planning data mining is essential.In this paper, we have considered the problem of data mining in the perspective of energy efficiency and good latency [12, 16].

2. DATA MINING FUNCTIONALITIES

Data mining functionalities include classification, clustering, association analysis, time series analysis, and outlier analysis.

- Classification is the process of finding a set of models or functions that describe and distinguish data classes or concepts, for the purpose of predicting the class of objects whose class label is unknown.
- Clustering analyzes data objects without consulting a known class model.
- Association analysis is the discovery of association rules displaying attribute-value conditions that frequently occur together in a given set of data.
- Time series analysis comprises methods and techniques for analyzing time series data in order to extract meaningful statistics and other characteristics of the data.
- Outlier analysis describes and models regularities or trends for objects whose behaviour changes over time.

3. DATA MINING APPLICATIONS

Data Mining in e-Commerce. Data mining enables the businesses to understand the patterns hidden inside past purchase transactions, thus helping in planning and launching new marketing campaigns in prompt and cost-effective way. e-commerce is one of the most prospective domains for data mining because data records, including customer data, product

data, users' action log data, are plentiful; IT team has enriched data mining skill and return on investment can be measured [2].

Data Mining in Industry. Data mining can highly benefit industries such as retail, banking, and telecommunications; classification and clustering can be applied to this area. One of the key success factors of insurance organizations and banks is the assessment of borrowers' creditworthiness in advance during the credit evaluation process. Credit scoring becomes more and more important and several data mining methods are applied for credit scoring problem.

On the enterprise side IoT offers services like Energy based applications such as operating management, as rigs and wells predictive maintenance, spill accident management, Smart healthcare applications, Smart retails, Smart agriculture applications, Smart banking applications, Smart building and Smart construction, Smart education , Smart insurance, Smart logistics and Smart manufacturing applications [4, 6].

Data Mining in Health Care. In health care, data mining is becoming increasingly popular, if not increasingly essential. Heterogeneous medical data have been generated in various health care organizations, including payers, medicine providers, pharmaceuticals information, prescription information, doctor's notes, or clinical records produced day by day. These quantitative data can be used to do clinical text mining, predictive modelling, survival analysis, patient similarity analysis, and clustering, to improve care treatment and reduce waste. In health care area, association analysis, clustering, and outlier analysis can be applied.

Data Mining in City Governance. In public service area, data mining can be used to discover public needs and improve service performance, decision making with automated systems to decrease risks, classification, clustering, and time series analysis which can be developed to solve this area problem.

Smart Cities provide the opportunity to integrate the physical infrastructures of the city such as the transportation sector, utilities, land, and city services [18]. The smart city model typically integrates the economic, social and environmental components of the city in a way that sustainably maximizes the efficiency of the city's primary systems.

Traffic Control - IoT devices or things such as smart phones, vehicle sensors, GPS are employed all over the city can served as data points such as time of traveling, incidence of heavy vehicles, accident prone zones and construction areas. Awareness of the causes behind traffic obstruction in the selected area is known with the help of these data points. To solve the traffic blockage problem we can use classification algorithm. The selected areas can be classified based upon the higher, lower or medium chances of traffic jam incidents in a specified location.

Suburban Electronic Meters - Traditional meters are replacing with smart electronic meters with a rapid pace, since smart meters can offer detailed description about real time energy consumption information in a digital way via e-mail or smart phones.

Pipeline Leak Detection – For municipal corporations, maintenance of water pipe leakages is burdensome more precisely with old pipes. Using outlier detection algorithm along with use of sensors, sound of water movement can be analyzed to spot leaks.

4. KEY ISSUES IN DATA MINING OF IOT

There are various issues involved in data mining in Internet of Things:

- Efficiency in data gathering - Energy efficiency, scalability and fault tolerance should be taken into consideration when data is to be collected from distributed sensor networks
- Data abstraction and aggregation - Managing massive data generated from IoT is a challenging task. Efficient mechanism should be adopted for data reduplication.
- Distributed data processing and mining - Due to nodes' constraints, paradigm shift is needed for prior level pre-processing of the data at each distributed nodes and an aggregated information is to be sent to sink node in order to optimize energy usage instead of sending all distributed data to server for processing [17].
- Data mining towards the next age of Internet - In an upcoming generations of Internet, latest trends and technologies like ubiquitous computing, semantic web, IPv6 technologies are going to be integrated with IoT. This will give rise to challenges for Data Mining due to heterogeneous unstructured data [1].

5. GREEN DATA MINING

Currently the majority of companies have data centers for storing their information; high-end server machines and devices are what data centers required [7]. There are two factors will be affected, a lot of power and deploying will be consumed and a lot of money will be consumed as well. However, applying green database techniques in Regional Health Authority (RHA) will helps to reduce the amount of energy that be consumed by data centers. Green database developing method is different with the developing method of normal database. The concepts of green database are derived from the green computing idea or green IT [11]. Green computing is a set of approaches that be designed to make data centers more efficient through reducing the consuming of power and cooling required. When it comes to compares the difference between the daily power consumption of typical data centers with the monthly power consumption of thousands homes, it will be noticed they are equivalent. The size of data center and the number of systems in data centers are the factors that will be used to identify the number of homes.

Green database is a type of database technique that used to manage power consumptions of data centers. The implementation of data warehouses and data centers requires patients and costs. Without a careful planning implementing a data warehouse can be very costly; especially when it comes to electricity. RHA has to look at the available possibilities to implementing a cost effective environmental friendly solution. Green database helps to save energy and save the environment [8]. The green includes the infrastructure, the optimization techniques and the virtualization solutions.

6. IOT WITH DATA MINING ARCHITECTURE

IoT gathers data from different places, again that data may contain required data for IoT aside [5]. Data collected by IoT is converted into serviceable information when we apply data mining to IoT, and then this information is converted into knowledge needed by user. The method of extracting patterns from the output of the data processing step and then

incorporating these patterns into the decision making process, which oversees the transformation of input into useful knowledge, all these steps are supervised by data mining process.

7. ENERGY CONSUMPTION IN CLOUD CENTER/DATA CENTER

Data centers are cost-effective infrastructures for storing large volumes of data and hosting large-scale service applications [9]. Data centers contain hundreds of thousands of servers, interconnected via switches, routers and high-speed links. Today, large companies such as Amazon, Google, Face book, and Yahoo! routinely use data centers for storage, web search, and large-scale computations. With the rise of cloud computing, service hosting in data centers has become a multi-billion dollar business that plays an important role in the future information Technology trade [10]. However, a large-scale computing infrastructure consumes huge amounts of electrical power leading to terribly high operational costs which will exceed the cost of the infrastructure in a very few years. In a datacenter, power consumption is mainly due to servers, networking devices, and cooling systems. There are two main approaches for reducing the energy consumption of datacenters: (a) shutting down devices or (b) scaling down performance. The former, commonly referred as Dynamic Power Management (DPM), results in the greatest savings, since the average workload often remains below 30% of its capacity in cloud computing systems.

Dynamic Voltage and Frequency Scaling (DVFS), a voltage reduction technique for battery-operated systems scaling, was introduced in the 90s, which dramatically reduces power consumption in large digital systems by adapting both voltage and frequency of the system with reference to dynamic workloads. Equipped with DVFS, a regulated system can adjust the supply voltage of a digital circuit at the functional boundary for the speed requirements, the temperature, and the technology parameters. DVFS describes the two power saving techniques (dynamic frequency scaling and dynamic voltage scaling) used to save power in embedded systems including cell phones. The servers left idle are put into sleep mode (DNS scheme) Dynamic Shutdown Scheme. According to Open Compute project report, 93% of the energy consumption in a data centre depends upon efficient utilization of computing resources at data centers [15].

8. ENERGY SAVING TECHNIQUES FOR SERVERS

The most important energy saving approaches for servers are;

- *Virtualization* - Virtualization is a prominent technique for reducing energy consumption in servers. It refers to making more than one Virtual Machine (VM) on a server. Using this technique decreases the number of hardware in use, improves the utilization of resources and reduces hardware and operating expenditure. Server virtualization also allows consolidation of server workloads. It may achieve energy saving by decreasing the amount of active and functioning servers with regards to Quality of Service requirements.
- *Dynamic Power Management*- Another important technology to decrease energy consumption is Dynamic Power Management. This technique is based on powering down the computing servers to save more energy. Putting inactive servers on the sleeping mode is the other way to reduce energy consumption in this technology. Moreover, joining virtualization technique with dynamic power management is another solution to save more energy for servers. This energy saving approach consolidates virtual machines on a subset of tangible servers and turned idle servers off or put them into sleep mode at the low utilization times. Then the servers which are powered off or putted into sleep mode are powered on when load increases and virtual machines from overloaded servers are moved to active servers that are ready to use with accessible resources.
- *The Dynamic Voltage/Frequency Scaling* - This approach sets the CPU power due to the presented load. By using this technique the power consumption will reduced, once CPU load is low.

9. ENERGY SAVING SOLUTIONS FOR NETWORK

Network infrastructure is the next main energy consumer in data centers. Around 30% of the entire energy usage which utilized for computation is consumed by data center network. Network in Data Centers includes switches and links. The utilization of a link is not directly balanced with its power consumption. Power consumption depends on capacity of the link instead of its utilization [13]. Energy usage of switches relies on a marketer and it is directed with the number of line cards and ports involved. It should be considered that power usage of both links and switches are important for reducing power consumption by network. The four following solutions are existed for saving energy in data center networks. Such as;

Adaptive Link - Adaptive Link Rate technique is extensively considered on wired networks. However, it has not been used on data center networks yet. This approach is based on the idea that energy usage of a link may be decreased its data rate, while traffic load of network links are low in most situations. In such cases, Adaptive Link Rate decreases link energy usage by dynamically setting link data rate to its utilization, while sleep mode method decreases energy usage through switching off network resources or placing them to sleep mode. Moreover, other active network resources should meet QoS requirements.

Virtual Network Embedding - Virtual network embedding is the next approach which is useful for reducing energy consumption of network. VNE is mostly useful while the network traffic is low. The aims of network virtualization is to use embedding algorithms to assign virtual network resources on a fewer number of physical infrastructure with an optimal approach.

Sleep Mode - The idle network resources could be switched off or put into sleep mode.

Green Routing - Recently, Energy aware routing or green routing for decreasing energy usage in data center network is studied. The main idea of Energy aware routing is to deliver routing service to less number of network resources to reduce energy usage, while sustaining network performance.

10. COMBINED ENERGY SAVING SOLUTIONS FOR SERVERS AND NETWORK

Recently, mixed approaches for saving energy in data centers are proposed extensively. 16 % of power usage may be saved by effecting network energy reducing approaches alone. It is worth mentioning that by combining server and network energy aware methods, energy could be save up 75 % of energy usage in data centers. By combining energy saving approach, network traffic

consolidation and server workload consolidation are cooperatively used. Three techniques for saving energy in data centers; such as Link state adaptation (LSA), Server load consolidation (NTC), Network traffic consolidation (SLC).

In LSA the power controller adapts to the state of links according to the information about traffic on each link. NTC decreases energy usage considerably by removing all redundancy in the network. This approach consolidates traffic on few numbers of links and switches, and idle links and switches are deactivated. SLC is a way to consolidate network traffic in a fewer links and switches to allow the controller to turn off unused resources. To achieve this goal, SLC transfers jobs to few numbers of servers to turn off unused servers. The results display that 16 % of energy could be stored by implementing only a link state adaptation approach, 75 % of energy could be stored by implementing both server load consolidation and network traffic consolidation together.

11. CONCLUSION

The idea of the Internet of Things derives from the need for all computers, instruments, and sensors in the world to be controlled, automated, and explored. Data mining technologies are combined with IoT technologies for decision-making support and device optimization in order to make wise decisions for both people and items in the IoT. Data mining includes the discovery of novel, fascinating and potentially useful data patterns and the application of algorithms to secret knowledge extraction. IoT produces huge quantities of useful heterogeneous data. Data mining systems are built in order to turn this data into information. The main purpose of the paper is to promote greenness, energy conservation and performance improvement in data mining tasks. An efficient and successful use of cloud computing tools can help to achieve Green Data Mining. The related research ideas concentrate primarily on energy-saving data centre approaches. More energy consumption is also minimised by choosing the best web services. However, due to the growing demand for data centre bandwidth and network access, data centre network and data centre server and network energy usage will rise rapidly in the future. This paper covered most of the technologies used in data centre data mining for energy saving during data mining.

REFERENCES

1. Feng Chen, Pan Deng et al., "Data Mining for the Internet of Things: Literature Review and Challenges", Hindawi Publishing Corporation, International Journal of Distributed Sensor Networks, Volume 2015, Article ID 431047, 14 pages.
2. G. Vamshi Krishna, "Data Mining Processes, Applications and Challenges for IoT", International Journal of Trend in Research and Development, Volume 4(3), ISSN: 2394-9333, May-Jun 2017, pp.157-160.
3. Sandhya Gundre, Ketaki Bhoyar et al., "Internet of Things Using Data Mining: Challenges And Applications", 2nd International Conference on Recent Developments in Science, Humanities & Management, ISBN : 978-93-87793-33-0, July 2018.
4. Priyanka Gupta, Rajan Gupta, "Data Mining Framework for IoT Applications", International Journal of Computer Applications (0975 – 8887), Volume 174 – No.2, September 2017, pp.4-7.
5. Akshat Savaliya, Aakash Bhatia et al., "Application of Data Mining Techniques in IoT: A Short Review", National Conference on Advanced Research Trends in Information and Computing Technologies (NCARTICT-2018),Volume 4, Issue 2, Print ISSN: 2395-1990, Online ISSN : 2394-4099, pp.218-223.
6. Furqan Alam, Rashid Mehmood et al., "Analysis of Eight Data Mining Algorithms for Smarter Internet of Things (IoT)", ELSEVIER, Science Direct, Procedia Computer Science 98 (2016) 437 – 442.
7. Hrishav Bakul Barua, Dr-Kartick Chandra Mondal, "Green Data Mining using Approximate Computing: An experimental analysis with Rule Mining", https://www.researchgate.net/ publication/329163375, sep 2018.
8. Omar S. Saleh, Samar H. Muhammad et al., "The impact of Green Database, Data Mining and Data Warehousing on the HealthCare System", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 7, July 2014, pp.101-105.
9. Chenlei Zhang, Abram Hindle, "A Green Miner's Dataset: Mining the Impact of Software Change on Energy Consumption", ACM, Copyright 2014.
10. IMRAN GHANI, NAGHMEH NIKNEJAD et al., "Energy Saving In Green Cloud Computing Data Centers: A Review", Journal of Theoretical and Applied Information Technology, Vol.74 No.1, E-ISSN: 1817-3195, 2015, pp.16-30.
11. Rushan Arshad, Saman Zahoor et al., "Green IoT: An Investigation on Energy Saving Practices for 2020 and Beyond", IEEE Access, Special Section on Future Networks: Architectures, Protocols, And Applications, Volume 5, 2017, pp. 15667-15681.
12. Yang Zhao, Chaobo Zhang et al., "A review of data mining technologies in building energy systems: Load prediction, pattern identification, fault detection and diagnosis", ELSEVIER, Energy and Built Environment 1 (2020) 149–164.
13. Yang Zhang, Tao Huang et al., "Big data analytics in smart grids: a review", Energy Informatics (2018) 1:8.
14. Priyank Sunhare, Rameez R et al., "Internet of things and data mining: An application oriented survey", Journal of King Saud University –Computer and Information Sciences, ELSEVIER, Production and hosting by Elsevier.
15. Maria Avgerinou, Paolo Bertoldi et al., "Trends in Data Centre Energy Consumption under the European Code of Conduct for Data Centre Energy Efficiency", Energies 2017, 10, 1470.
16. Korhan Cengiz and Tamer Dag, "A review on the recent energy-efficient approaches for the Internet protocol stack", Cengiz and Dag EURASIP Journal onWireless Communications and Networking (2015) 2015:108.
17. Yanik Ngoko, Alfredo Goldman et al., "Service selection in web service compositions optimizing energy consumption and service response time", Journal of Internet Services and Applications 2013, 4:19.
18. Hisham Elhoseny, Mohamed Elhoseny et al., "A Framework for Big Data Analysis in Smart Cities", Springer International Publishing AG, part of Springer Nature 2018, A. E. Hassanien et al. (Eds.): AMLTA 2018, AISC 723, pp. 405–414, 2018.