Optimizing IoT Data Transmission in Smart Agriculture: A Comparative Study of Reduction Techniques

2023 5th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) | 979-8-3503-3752-5/23/$31.00 ©2023 IEEE | DOI: 10.1109/HORA58378.2023.10156757

Raad.M.KHALEEFAH,
Department of Computer Science, Biotechnology and Environmental Centre
Fallujah university
Fallujah, Iraq
email: rdaljwahry@gmail.com

Nadhim A. M Al-isawi
Department of Computer Science, College of Medicine
Fallujah university
Fallujah, Iraq
email mm.nadhimabid@gmail.com

M. K. Hussein
Dept. Electronics, Telecommunication & Information Technology
University Politehnica of Bucharest
Bucharest, Romania
email: M.k.hussein93@gmail.com

N.A.M.Alduais
Faculty of Computer Science and Information Technology
Universiti Tun Hussein Onn Malaysia
Johor, Malaysia
email :nayef@uthm.edu.my

***Abstract*— Smart agriculture, driven by Internet of Things (IoT) technologies, has revolutionized traditional farming practices. The integration of sensors and devices in the agricultural environment has generated a vast amount of data for analysis and decision-making. However, managing and transferring this abundant data poses challenges, particularly in terms of data transfer and storage. Therefore, the purpose of this study is to investigate and compare different data reduction techniques for optimizing IoT data transmission in smart agriculture. These techniques aim to reduce the size or quantity of transmitted data samples, leading to reduced energy consumption, improved network efficiency, and minimized storage and processing requirements. The selected data reduction techniques include sampling, quantization, and deduplication. The evaluation of these techniques was conducted using a real-time dataset specifically designed for smart agriculture systems. The dataset comprises five features: ambient air temperature, ambient air moisture, soil temperature, soil moisture, and soil pH. The comparison of the techniques was based on three metrics: information loss, computation cost, and energy consumption, taking into account the size of each transmitted sample. Python programming language was employed to implement the three data reduction techniques. The results revealed that there exists a trade-off among information loss, computational cost, and energy consumption, which necessitates careful consideration when selecting a data reduction technique for IoT nodes in smart agriculture applications. This analysis provides valuable insights for making informed decisions that strike a balance between these factors, considering the specific requirements and constraints of each application.**

Keywords—IoT, Smart agriculture, Data Transmission

I. INTRODUCTION AND RELATED WORKS

Food and nutrition are the essential needs for all living things, including humans. Agriculture is the activity of growing food conducted by humans, thus is the fundamental activity in human society, allowing not only self-sustaining life, but also become the base to other activities such as economics and technological improvement. In modern society, as the need for food all over the world increases, while the food security caused by unstable climates and environments, and man-made incidents keeps on declining [1], it is important for agriculture to be improved and upgraded as well, so it can solve the increasing need for food. One of the modern solutions that can be deployed relatively easily but still able to provide invaluable improvements in agriculture soon is using IoT and smart technology in agriculture. IoT and smart technology often mean the same context and technology, which is the connection and network between physical objects such as instruments, vehicles, electronics, devices and any items with embedded system of electronics, circuits, software, sensors and network connectivity, so they could communicate and interact with each other, thus realizing automation, remote controlling and management, and improved accuracy and efficiency in most systems [2].

Efficient data transmission is crucial for IoT-based smart farming systems. Transmitting large amounts of raw sensor data from IoT nodes to base stations can lead to issues such as network congestion, increased energy consumption, and limited network capacity [3][4]. To address these challenges, it is crucial to explore data reduction techniques that can optimize data transfer while maintaining data integrity[5].

IoT based Water Quality Control System for Aquaculture: Tran Duc Chuyen et al proposed a project in Vietnam that focuses on aquaculture [6]. Their system includes wirelessly connected sensor groups, an IoT control circuit, a solar panel for power, a cloud server for data transfer, and web and mobile applications. By monitoring water environmental indicators in shrimp ponds, the system aims to improve the water quality of the shrimp pools.

Mohamed et al conducted research to showcase the application of blockchain and IoT technologies in agriculture [7]. This research emphasizes the decentralization of computation and management processes, addressing issues such as security, resource consumption, privacy, and decision-making within existing IoT systems, including those used in agriculture.

Yu-Hsin et al proposed a project that demonstrates the application of cloud computing, IoT, and PM technology in smart agriculture [8]. The system involves setting up sensors and cameras to observe plants and soil conditions. By analyzing the collected data, the system helps control and monitor plants, ultimately improving crop quality and quantity.

979-8-3503-3752-5/23/$31.00 ©2023 IEEE

Pavan Patil et al worked on upgrading AutoGrow, an AI and ML-based system for precision agriculture [9]. They designed a data logging and sensing sub-system using IoT technology to further enhance the system's efficiency in acquiring and optimizing resource usage in agriculture.

Micah Bogdanoff et al proposed an automated, inexpensive, and user-friendly system for agriculture [10]. Their system utilizes sensors and actuators to monitor soil moisture content and control water valves in multiple irrigation zones. Communication is facilitated through cellular networks, connecting to a central communication node and allowing users to interact with the system through a cloud server.

Rolf A. Kjellby et al proposed a long-range, self-powered IoT device for precision agriculture and aquaponics [11]. Their system monitors environmental parameters such as humidity and pH, sending the data to cloud servers for analysis. The analyzed data helps optimize resource management in agriculture activities.

Paul Vandôme et al developed a system for monitoring drip irrigation in Tunisia [12]. Their system utilizes wireless sensor networks (WSN) with low resource consumption. By providing real-time data, the system facilitates decision-making and improves water resource management in agricultural activities. TABLE I shows the summarization of the related work.

TABLE I: SUMMARIATION OF THE RELATED WORK

Ref	Prototype	Sensors	Connectivity	Architecture	Energy Consumption
[6]	Arduino Uno R3	Temperature, pH, Dissolved Oxygen (DO)	Wi-Fi	WSN, IoT	N/A
[7]	Mobile, Web & PDA Applications	pH, Temperature, Humidity, Soil Moisture	Wi-Fi, Bluetooth	Blockchain, IoT	N/A
[8]	Blynk IoT Platform, Web and Mobile Application	Humidity	Wi-Fi	IoT, Cloud Computing, Preventive Maintenance (PM)	N/A
[9]	Raspberry-Pi-3 & Arduino UNO	Soil Moisture, Nitrogen, Phosphorus, Potassium, pH	Wi-Fi	IoT	N/A
[10]	ESP8266	Soil Moisture, Temperature	2.4GHz LPWAN	WSN, IoT	N/A
[11]	nRF52840 microcontroller	Temperature, Humidity, Visible Light, Battery Level, Air Pressure, pH, Soil Moisture etc.	Bluetooth 5 with Long Range Support	WSN, IoT	N/A
[12]	Arduino Pro-Mini	Soil Moisture	LoRa WAN	WSN, IoT	N/A

In the following subsection, we discuss some data reduction methods.

A. Sampling

Sampling is a data reduction technique that involves selecting a subset of data points from a larger dataset [13] . It is widely used in various domains to reduce the amount of data without losing critical information. In the context of IoT data, sampling techniques aim to capture representative data points while reducing the overall volume of transmitted data. Sampling methods can vary, including random sampling, stratified sampling, or time-based sampling [14]. Each method has its own advantages and considerations depending on the application requirements. Random sampling provides an unbiased representation of the data, while stratified sampling ensures proportional representation of different classes or categories within the data. Sampling allows for reducing the data size, which in turn reduces storage requirements and transmission costs. However, the selection of appropriate sampling rates and strategies is crucial to strike a balance between data reduction and preserving essential information.

B. Quantization

Quantization is a process of reducing the precision or granularity of numerical data[15]. It involves mapping continuous or high-resolution data values to a smaller set of discrete values. The purpose of quantization is to decrease data complexity and storage requirements while retaining the general trends and patterns in the data. In IoT applications, quantization is commonly used to represent sensor readings with a limited number of levels or categories[16]. The selection of appropriate quantization intervals or thresholds depends on the data characteristics and the specific application requirements. Adaptive quantization techniques, such as Lloyd-Max quantization, adjust the quantization levels dynamically based on the data distribution. Quantization reduces the amount of transmitted data by representing it with a smaller number of bits or symbols. However, it introduces a trade-off between data resolution and information loss. The choice of quantization levels should be carefully determined to minimize loss while achieving sufficient data reduction.

C. Deduplication

Deduplication, also known as data deduplication or duplicate detection, is a technique used to identify and eliminate redundant or duplicate data entries[17]. It aims to reduce data redundancy and storage overhead by storing only unique instances of data. In IoT applications, deduplication is employed to eliminate repeated sensor readings or duplicate data entries resulting from transmission errors or multiple sensors capturing the same information. Various deduplication algorithms exist, such as hashing-based methods, content-aware techniques, or delta encoding. Deduplication reduces the amount of data to be stored or transmitted by eliminating duplicate entries. However, it introduces a challenge in scenarios where data changes frequently, as frequent updates may result in increased computational overhead for detecting duplicates.

II. IMPLEMENTATION AND RESULTS

In this section, we present the implementation details of the used methods for data reduction in smart agriculture applications. We describe the steps involved in each method, including sampling, quantization, and deduplication in the following pseudo code. Furthermore, we provide an analysis of the results obtained from applying these methods to a given dataset.

Algorithms*: // Sampling / *Quantization / Deduplication//

```
1.  Input: D , f, q:

2.  Output S, Q, D', L_s, L_q, L_d, C_s, C_q, C_d,
    E_s, E_q, E_d.

3.  Begin:
4.  // Sampling //
5.  N = f * len(D)
6.  S = D.sample(n=N, random_state=1)
7.  L_s = compute_loss(D, S)
8.  C_s = compute_computation_cost(S)
9.  E_s = compute_energy_consumption(S)

10. // Quantization //
11. T = QuantileTransformer(n_quantiles=q,
    random_state=0)
12. T.fit(D)
13. Q = T.transform(D)
14. Q = pd.DataFrame(Q, columns=D.columns)
15. L_q = compute_loss(D, Q)
16. C_q = compute_computation_cost(Q)
17. E_q = compute_energy_consumption(Q)

18. // Deduplication //
19. D' = D.drop_duplicates()
20. L_d = compute_loss(D, D')
21. C_d = compute_computation_cost(D')
22. E_d = compute_energy_consumption(D')
23. End
```

The symbols used in the pseudo code are as follows:
D: Original dataset
f: Sampling fraction
q: Number of quantiles
S: Sampled dataset
Q: Quantized dataset
D': Deduplicated dataset
L_s: Information loss due to sampling
L_q: Information loss due to quantization
L_d: Information loss due to deduplication
C_s: Computation cost of sampling
C_q: Computation cost of quantization
C_d: Computation cost of deduplication
E_s: Energy consumption of sampling
E_q: Energy consumption of quantization
E_d: Energy consumption of deduplication

We analyzed the performance of three data reduction techniques - Sampling, Quantization, and Deduplication - on a Smart Agriculture dataset. This dataset consists of five features: ambient air temperature, ambient air moisture, soil temperature, soil moisture, and soil pH. The metrics used for comparison were information loss (measured in data units), computation cost (measured in computational cycles), and energy consumption (measured in millijoules) based on the size of each transmitted sample. TABLE II presents the statistical analysis of the Smart-Agriculture dataset, including the count, mean, standard deviation, minimum, 25th percentile (Q1), median (50th percentile), 75th percentile (Q3), and maximum values for each feature. The features analyzed include ambient air temperature, ambient air moisture, soil temperature, soil moisture, and soil pH. These statistics provide valuable insights into the characteristics and variability of the dataset, aiding in understanding the environmental conditions in smart agriculture applications.

TABLE II: STATISTICAL ANALYSIS OF USED DATASET [18]

FEATURE	COUNT	MEAN	STD	MIN	25%	50%	75%	MAX
Ambient Air Temp	499590	21.40	6.81155	0	22	23	24	30
Ambient Air Moisture	499590	68.6623	23.9122	0	64	76	84	99
Soil Temperature	499590	23.7031	1.44406	21	23	23	25	32
Soil Moisture	499590	61.8741	11.6712	29	57	61	65	97
Soil Ph	499590	4.84829	2.107	-10	4	6	6	17

A. Information Loss

From the results, we found that Deduplication led to the highest information loss (2.42 x 10^6 data units), followed by Sampling (1.25 x 10^6 data units), and then Quantization (1.11 x 10^3 data units). Deduplication removes all redundant data, which in the context of Smart Agriculture could mean removing multiple identical sensor readings of ambient conditions or soil parameters. Although this significantly reduces the data volume, it could also result in the loss of important information, particularly if certain readings are consistently similar but still crucial for analysis.

Sampling caused significant information loss as it involves selecting only a subset of the data. Depending on the frequency of sampling, potentially important readings (such as abrupt changes in soil moisture or pH) might have been missed.

Quantization, which caused the least information loss, reduced the complexity of the data by decreasing the number of unique values while preserving the overall trends and patterns in the sensor readings.

B. Computation Cost

In terms of computation cost, Quantization was the most computationally expensive method (19,983,728 computational cycles), followed by Sampling (11,990,160 computational cycles), with Deduplication being the least costly (707,280 computational cycles). The complex computations involved in mapping original data values to a smaller set of values in Quantization likely contributed to this. Deduplication, which involves identifying and removing duplicate values, was the least computationally expensive. However, in scenarios where the data changes frequently, such as sensor data from rapidly changing environmental conditions, the computational cost of Deduplication could potentially increase.

C. Energy Consumption

Energy consumption was directly proportional to the computation cost for each method. Consequently, Quantization consumed the most energy (199,837.28 millijoules), while Deduplication consumed the least (7,072.80 millijoules). For IoT nodes in Smart Agriculture, where energy efficiency is a critical factor, this could be a deciding factor in choosing the appropriate data reduction technique.

D. Implications for IoT nodes in Smart Agriculture

These techniques could be highly beneficial in the context of Smart Agriculture. IoT nodes, such as sensors for ambient air temperature, ambient air moisture, soil temperature, soil moisture, and soil pH, can use these methods to reduce the amount of data that needs to be stored and transmitted. This results in savings in storage, computational power, and energy. However, the choice of method would depend on the specific requirements of the application. If the application can tolerate some loss of information for lower energy consumption, then Deduplication or Sampling might be preferred. If maintaining data integrity is crucial, then Quantization might be a better choice, despite its higher computational cost and energy consumption. The visual representation of the results in the figures Fig.1, Fig.2, and Fig.3 clearly shows the differences in the metrics for each of the data reduction techniques. The figure for Information Loss shows a significant difference in the amount of information lost between Quantization and the other two methods. The Computation Cost and Energy Consumption figures mirror each other due to the direct correlation between these two metrics.

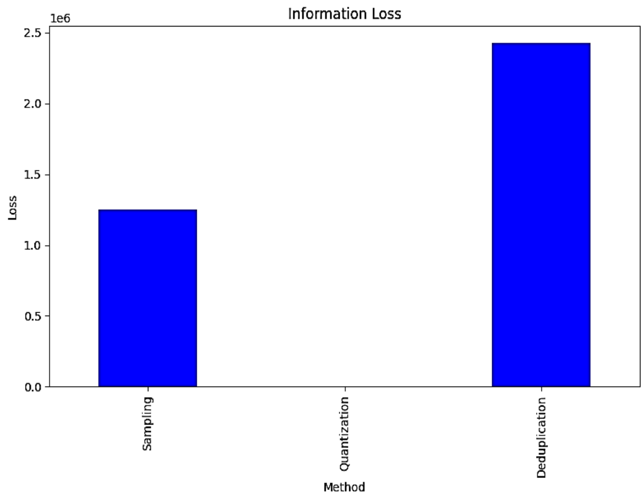

Fig.1 Results of information loss metric for applying different methods

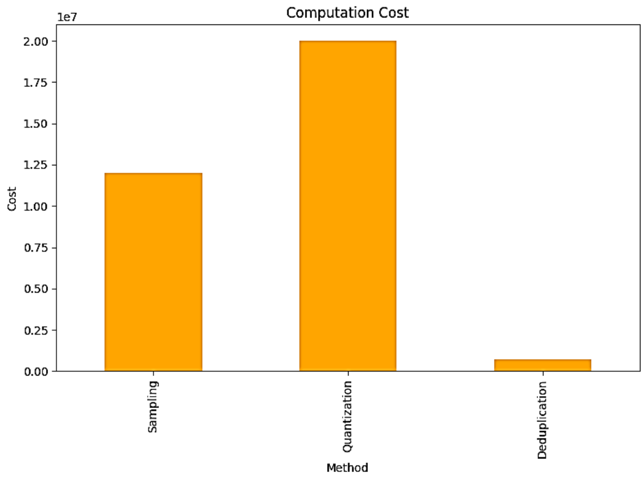

Fig.1 Results of computation cost metric for applying different methods

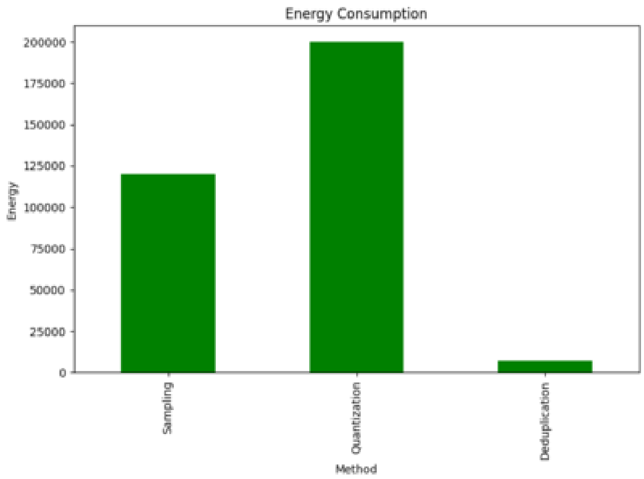

Fig.1 Results of energy consumption metric for applying different methods

III. CONCLUSIONS

The purpose of this study is to investigate and compare different data reduction techniques for optimizing IoT data transmission in smart agriculture. These techniques aim to reduce the size or quantity of transmitted data samples, resulting in reduced energy consumption, improved network efficiency, and minimized storage and processing requirements. The selected data reduction techniques include sampling, quantization, and deduplication. The evaluation of these techniques was conducted using a real-time dataset specifically designed for smart agriculture systems. The dataset comprises five features: ambient air temperature, ambient air moisture, soil temperature, soil moisture, and soil pH. The comparison of the techniques was based on three metrics: information loss, computation cost, and energy consumption, taking into account the size of each transmitted sample. The Python programming language was employed to implement the three data reduction techniques. The results revealed a trade-off between information loss, computational cost, and energy consumption, highlighting the need for careful consideration when selecting a data reduction technique for IoT nodes in smart agriculture applications.

This analysis provides valuable insights for making informed decisions that strike a balance between these factors, considering the specific requirements and constraints of each application.

REFERENCES

[1] Ghojogh, Benyamin, and Mark Crowley. "Principal sample analysis for data reduction." 2018 IEEE International Conference on Big Knowledge (ICBK). IEEE, 2018.

[2] Tutmez, Bulent. "Robust learning algorithm based on agreement among soil sampling techniques." Applied Soft Computing 137 (2023): 110123.

[3] Alduais, N. A. M., Abdullah, J., & Jamil, A. (2018). APRS: adaptive real-time payload data reduction scheme for IoT/WSN sensor board with multivariate sensors. International Journal of Sensor Networks, 28(4), 211-229.

[4] Abdullah, Jiwa, M. K. Hussien, N. A. M. Alduais, M. I. Husni, and Ansar Jamil. "Data reduction algorithms based on computational intelligence for wireless sensor networks applications." In 2019 IEEE 9th Symposium on Computer Applications & Industrial Electronics (ISCAIE), pp. 166-171. IEEE, 2019.

[5] Hussein, M. K., Marghescu, I., & Alduais, N. A. (2022). Performance of Data Reduction Algorithms for Wireless Sensor Network (WSN) using Different Real-Time Datasets: Analysis Study. International Journal of Advanced Computer Science and Applications, 13(1).

[6] Gray, Robert M., and David L. Neuhoff. "Quantization." IEEE transactions on information theory 44.6 (1998): 2325-2383.

[7] Azar, Joseph, et al. "Robust IoT time series classification with data compression and deep learning." Neurocomputing 398 (2020): 222-234.

[8] He, Qinlu, Zhanhuai Li, and Xiao Zhang. "Data deduplication techniques." 2010 international conference on future information technology and management engineering. Vol. 1. IEEE, 2010.

[9] Josephine Mbandi, May 3, 2021, "Soil Data Collection using Wireless Sensor Networks (WSN) and Offsite Visualization: Case-study of the Innovative Solutions for Digital Agriculture (iSDA) Project in Kenya", IEEE Dataport, doi: https://dx.doi.org/10.21227/v2m5-qs54.

[10] Quy, V. K., Hau, N. V., Anh, D. V., Quy, N. M., Ban, N. T., Lanza, S., Randazzo, G., & Muzirafuti, A. (2022). IOT-enabled Smart Agriculture: Architecture, applications, and challenges. Applied Sciences, 12(7), 3396.

[11] Gokhale, P., Bhat, O., & Bhat, S. (2018). Introduction to IOT. International Advanced Research Journal in Science, Engineering and Technology, 5(1), 41-44.

[12] Chuyen, T. D., Nguyen, D. D., Cuong, N. C., & Thong, V. V. (2023). Design and manufacture control system for water quality based on IOT technology for aquaculture in the Vietnam. Bulletin of Electrical Engineering and Informatics, 12(4), 1893–1900.

[13] Kassim, M. R. (2022). Applications of IOT and Blockchain in smart agriculture: Architectures and challenges. 2022 IEEE International Conference on Computing (ICOCO).

[14] Hung, Y.-H. (2019). Developing an agriculture internet of things (IOT) application taking the plants humidity preventive maintenance (PM) as an example. 2019 2nd World Symposium on Communication Engineering (WSCE).

[15] Patil, P., Kestur, R., Rao, M., & C†, A. (2023). IOT based data sensing system for autogrow, an autonomous greenhouse system for Precision Agriculture. 2023 IEEE Applied Sensing Conference (APSCON).

[16] Bogdanoff, M., & Tayeb, S. (2020). An ISM-Band Automated Irrigation System for Agriculture IoT. In 2020 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS).

[17] Kjellby, R. A., Cenkeramaddi, L. R., Froytlog, A., Lozano, B. B., J, S., & Bhange, M. (2019). Long-range & Self-powered IoT Devices for Agriculture & Aquaponics Based on Multi-hop Topology.

[18] Vandôme, P., Leauthaud, C., Moinard, S., Sainlez, O., Mekki, I., Zairi, A., & Belaud, G. (2023). Making technological innovations accessible to agricultural water management: Design of a low-cost wireless sensor network for drip irrigation monitoring in Tunisia. Smart Agricultural Technology, 4, 100227.