Extreme Gradient Boosting Model-Based Optimal Configuration Of Iot Network Using Lora Transceiver

Namnsowo Edet Akpan[1]
Department of Electrical and Electronics Engineering
University of Uyo, Uyo
namnsowoakpan@gmail.com

Akaninyene Bernard Obot[2]
Department of Electrical and Electronics Engineering
University of Uyo, Uyo
akaninyeneobot@uniuyo.edu.ng

Udofia, Kufre M[3]
Department of Electrical and Electronics Engineering
University of Uyo, Uyo
kufremudofia@uniuyo.edu.ng

***Abstract*— The Extreme Gradient Boosting (XGBoost) model-based optimal configuration of IoT network using LoRa transceiver is presented. The work is focused on addressing the challenges of selecting the parameters values combination that will minimize the energy needed to successfully transmit each data packet. The case study 3000 records dataset with 10 features was generated by simulating an IoT network based on Semtech SX1276 transceiver operating at 868 MHz. The data was preprocessed and the energy efficiency computed as PDR/energy is normalized with the MinMax approach with minimum value of 0% and maximum value of 100%. The XGBoost was then trained and validated using 75% by 25 % data splitting ratio. The results showed a trend of increased energy consumption with increasing distance giving 0.8193 mJ of energy consumption 521.13 m distance and 3.3220 mJ of energy consumption at 886.98 m; aligning with expectations in low-power wide-area networks. Also, 95% packet delivery ratio (PDR) is achieved at TD = 10.85 m, which is consistent with strong signal conditions and minimal path loss. The time on air (ToA) is minimal (0.020 ms) at TD = 210.30 m, due to the low payload size and moderate bandwidth, while longer distances and larger payloads predictably result in higher ToA values (such as 0.215 ms at TD = 886.98 m). In all, the study observed the strength and weakness of the XGBoost model in this application and recommends that further analysis is required to address the weakness to make the XGBoost model more accurate in the prediction of the optimal configuration values for the IoT network.**

Keywords*— *IoT Network, Extreme Gradient Boosting Model, Optimal Configuration, Sensor Node, LoRa Transceiver , Energy Consumption Model

1. Introduction

In recent years, LoRa-based Internet of Things (IoT) networks have gained wide application [1,2,3]. The long range and low power demand feature of the LoRa transceiver has made it possible to access long distance, even enabling direct earth to satellite communication [4,5]. These capabilities are due to the different parameter configurations afforded by the Lora technologies.

LoRa transceivers have different spreading factors, bandwidth, payload size and transmission power which affect the energy efficiency and transmission range [6,7]. Also, parameters like the duty cycle and coding rate also affect the packet delivery ratio and energy efficiency [8]. In operation, careful selection of the different parameter settings is required to ensure energy efficiency. This is particularly important in IoT sensor nodes which are in many cases battery-powered with finite battery lifespan that is dependent on the energy consumption of the sensor node [9,10,11]. Accordingly, this work major aim is the application of XGBoost model to predict the parameter settings that will afford the lowest energy consumption for each packet that is successfully delivered over a given distance within the IoT network coverage area [12,13]. The study is based on simulated dataset of an IoT based on LoRaWAN.

2. Methodology: development of the Extreme Gradient Boosting Model for LoRa Optimal Parameter Configuration

In this work Extreme Gradient Boosting (XGBoost) is employed to select appropriate combinations of LoRa transceiver parameters values for optimal energy consumption. The XGBoost model is well-suited for structured data problems involving non-linear, high-dimensional, and sparse input features making it a strong candidate for predicting energy efficiency metrics in LoRaWAN systems. XGBoost can model complex relationships between parameters like Spreading Factor (SF), Transmission Power (P_Tx), Bandwidth (BW), Time on Air (ToA), and Transmission Distance (TD), all of which directly or indirectly affect energy consumption in IoT communication. Given a dataset $D = \{(x_i, y_i)\}_{i=1}^{n}$ where, $x_i \in R^d$ are the transmission parameters and $y_i \in R$ is the target variable (Efficiency), XBGoost learns a function:

$$\hat{y}_i^{(t)} = \sum_{k=1}^{K} f_k(x_i), \quad f_k \in \mathcal{F} \tag{1}$$

Where, $\mathcal{F}$ is a space of regression trees, f_k is a individual regression tree in the ensemble, and K is the number of boosting iterations. The learning objective in iteration t is:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{n} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) + \Omega(f_t) \tag{2}$$

Where, l is used to denote the convex loss function, while the Ω is used to denote the penalizing complexity for the regularization term.

$$\Omega(f) = \gamma T + \frac{1}{2}\lambda \sum_{j=1}^{T} w_j^2 \tag{3}$$

The summary of the case study dataset generated by simulating an IoT network based on Semtech SX1276 transceiver operating at 868 MHz is show Table 1 while Table 2 shows the correlation matrix of the parameters. Notably, there are 10 features in the dataset consisting of 3000 records, as shown in Table 1. According to the correlation matrix, the transmission distance, TD is highly correlated with the transmitter power (P_Tx) and the packet delivery ratio (PDR). Td is also correlated fairly high with the energy consumption. In essence, while, higher TD give rise to high P_Tx and lower PDR due to the negative correlation, the energy consumption may increase depending on the contributions of some other parameters. This is why the correlation between TD and energy is not as high as the value between TD and P_Tx.

Table 1 The parameters in the case study dataset

Column Name	Non-Null Count	Data Type	Description
Node ID	3000	int64	Unique identifier for each simulated sensor node in the network.
TD	3000	float64	Transmission distance from node to gateway, measured in meters.
SF	3000	int64	Spreading factor, ranging from 7 to 12, influencing time-on-air and range.
BW	3000	float64	Transmission bandwidth, typically 125 kHz, 250 kHz, or 500 kHz.
CR	3000	int64	Coding rate index; represents LoRa error correction capability.
DC	3000	float64	Duty cycle; fraction of active time over a transmission cycle.
Payload Size	3000	int64	Size of transmitted data packet in bytes.
PDR	3000	float64	Packet Delivery Ratio; probability of successful delivery (0 to 1).
P_Tx (mW)	3000	float64	Transmission power in milliwatts, based on distance and signal parameters.
ToA (s)	3000	float64	Time-on-air; duration required to transmit a full packet.
Energy (mJ)	3000	float64	Total energy consumed per transmission event, computed in millijoules.

Table 2 The correlation matrix for the features in the dataset

	TD	SF	BW	CR	DC	PayloadSize	PDR	P_Tx (mW)	ToA (s)	Energy (mJ)
TD	1.000	-0.014	-0.009	0.023	0.005	-0.024	-0.980	0.972	-0.005	0.637
SF	-0.014	1.000	-0.040	0.017	0.000	-0.008	-0.186	-0.004	-0.223	-0.140
BW	-0.009	-0.040	1.000	-0.008	-0.001	0.006	0.017	-0.013	-0.647	-0.373
CR	0.023	0.017	-0.008	1.000	0.024	0.013	-0.026	0.024	0.220	0.144
DC	0.005	0.000	-0.001	0.024	1.000	-0.013	-0.005	0.005	-0.001	-0.001
PayloadSize	-0.024	-0.008	0.006	0.013	-0.013	1.000	0.026	-0.022	0.521	0.277
PDR	-0.980	-0.186	0.017	-0.026	-0.005	0.026	1.000	-0.954	0.050	-0.597
P_Tx (mW)	0.972	-0.004	-0.013	0.024	0.005	-0.022	-0.954	1.000	-0.002	0.656
ToA (s)	-0.005	-0.223	-0.647	0.220	-0.001	0.521	0.050	-0.002	1.000	0.566
Energy (mJ)	0.637	-0.140	-0.373	0.144	-0.001	0.277	-0.597	0.656	0.566	1.000

3. Results and discussion

The data was preprocessed and the energy efficiency computed as PDR/energy is normalized with the MinMax approach with minimum value of 0% and maximum value of 100%. The XGBoost was then trained and validated using 75% by 25 % data splitting ratio. The XGBoost was trained select the parameter value combinations that give the highest energy efficiency for any given transmission distance, TD. The graph in Figure 1 presents the energy predictions from the baseline XGBoost model across different transmission distances (TD). While the model performs reasonably in predicting energy consumption trends, some predictions such as negative energy values at TD = 10.85 m (-0.1507 mJ) and TD = 210.30 m (-1.1359 mJ) are physically unrealistic. These values likely reflect model over fitting or extrapolation errors. For valid positive predictions, a trend of increased energy consumption with increasing distance is evident (e.g., 3.3220 mJ at 886.98 m), aligning with expectations in low-power wide-area networks (LPWANs).

Again, as seen in Figure 2, the XGBoost model associates higher packet delivery ratio (PDR) with shorter transmission distances and optimal parameter configurations. For instance, 95% PDR is achieved at TD = 10.85 m, consistent with strong signal conditions and minimal path loss. However, the model fails to maintain robust delivery as distance increases, with PDR dropping to 7% at 886.98 m, despite a relatively high SF and bandwidth.

The results in Table 3 and Figure 4.28 outline the time required to transmit a packet under optimal settings. The time on air (ToA) values are mostly consistent with the theoretical relationships between SF, payload, and bandwidth. For example, ToA is minimal (0.020 ms) at TD = 210.30 m, due to the low payload size and moderate bandwidth, despite the use of SF = 11. In contrast, longer distances and larger payloads predictably result in higher ToA values (e.g., 0.215 ms at TD = 886.98 m). This aligns well with expectations and suggests that XGBoost does capture the structural dependencies that define ToA. However, since ToA does not directly incorporate network congestion or real-world delays, the model's performance on this metric is informative but not comprehensive.

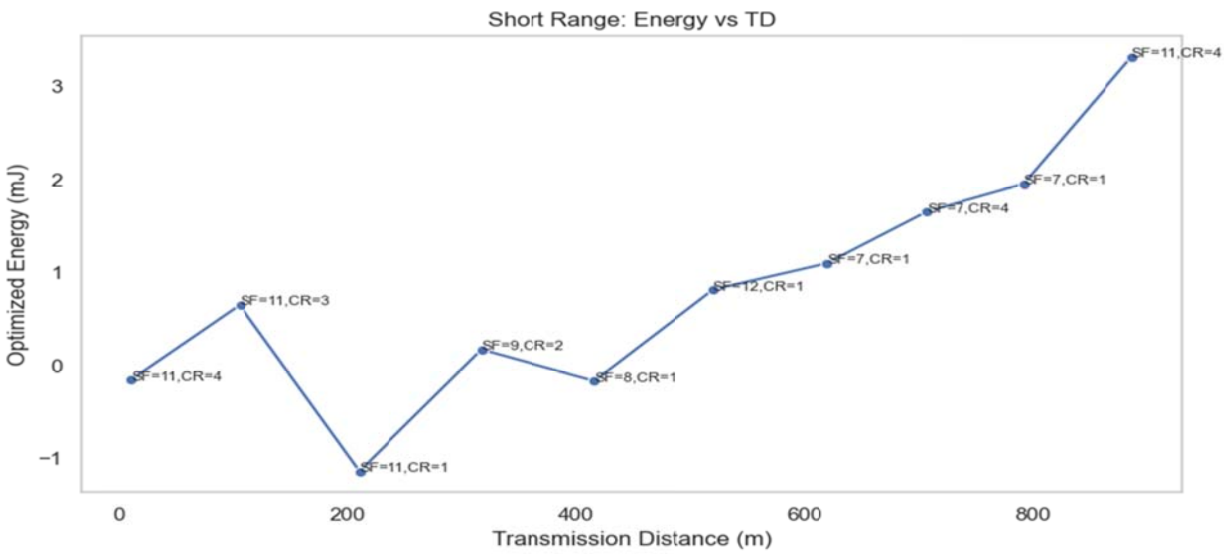

Figure 1: Optimal parameter configuration energy optimization

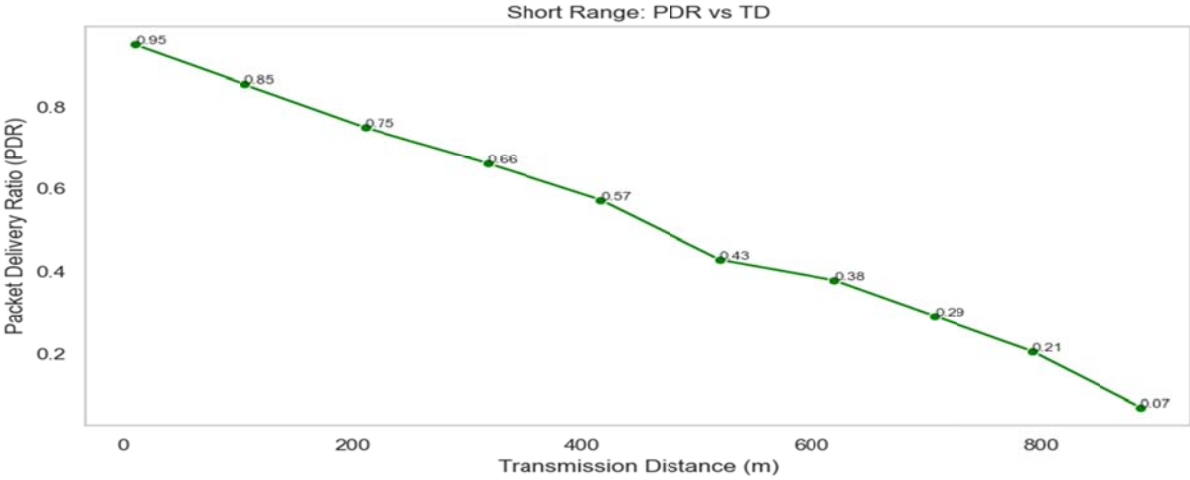

Figure 2: Packet delivery ratio for optimal configuration

Figure 3: Time-on-air for optimal configuration

The SNR values in the graph plot of Figure 4 show a well-structured decline as transmission distance increases, starting from 97.59 dB at TD = 10.85 m to 52.57 dB at TD = 886.98 m. This decline mirrors theoretical expectations derived from the logarithmic decay in received power as defined by the Friis transmission equation and path loss models. The values further suggest that while the XGBoost model may not optimize reliability directly, it indirectly aligns with physical-layer dynamics, capturing distance-related signal attenuation accurately. Such consistent SNR behavior provides a stable foundation for subsequent optimization frameworks.

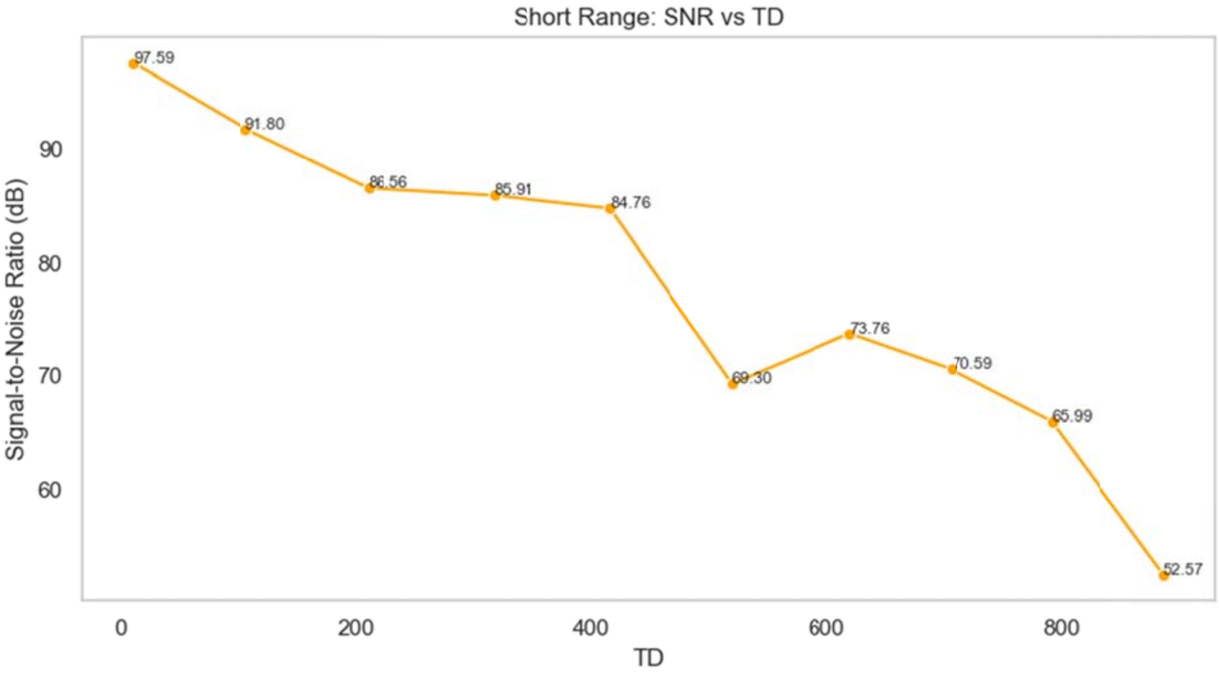

Figure 4: SNR for optimal configuration

In all, the XGBoost model exhibits strong representational capability for patterns such as ToA and SNR. However, its tendency to produce negative or non-physical energy values renders it unfit for direct deployment without correction mechanisms. Its limitations in optimizing energy-PDR trade-offs and lifetime projections further stress the need for enhanced meta-optimization techniques to guide XGBoost-based predictions toward feasible, efficient, and reliable transmission configurations in LoRa networks.

4. Conclusion

The approach for selecting the best parameter configurations at any given transmission distance for maximum energy efficiency is presented. The approach rely on the Extreme Gradient Boosting (XGBoost) model which is trained with simulated dataset of the case study IoT network. The IoT network studied used the LoRa technology for its communication link. In this case, the best combination of the LoRa transceiver parameters and the IoT network that will yield the smallest energy per delivered packet is predicted by the XGBoost model. In all, the study observed the strength and weakness of the XGBoost model in this application and recommends that further analysis is required to address the weakness to make the XGBoost model more accurate in the prediction of the optimal configuration values for the IoT network.

References

1. Liaqat, R. M., Branch, P., & But, J. (2022, November). Lora based linear network applications, design considerations and open challenges: A review. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems* (pp. 913-917).
2. Yao, F., Ding, Y., Hong, S., & Yang, S. H. (2022). A survey on evolved LoRa-based communication technologies for emerging internet of things applications. *International Journal of Network Dynamics and Intelligence*, 4-19.
3. Zourmand, A., Hing, A. L. K., Hung, C. W., & AbdulRehman, M. (2019, June). Internet of things (IoT) using LoRa technology. In *2019 IEEE international conference on automatic control and intelligent systems (I2CACIS)* (pp. 324-330). IEEE.
4. Ullah, M. A., Pasolini, G., Mikhaylov, K., & Alves, H. (2023). Understanding the limits of LoRa direct-to-satellite: The Doppler perspectives. *IEEE Open Journal of the Communications Society*, *5*, 51-63.
5. Li, C., & Cao, Z. (2022). Lora networking techniques for large-scale and long-term iot: A down-to-top survey. *ACM Computing Surveys (CSUR)*, *55*(3), 1-36.
6. Gkotsiopoulos, P., Zorbas, D., & Douligeris, C. (2021). Performance determinants in LoRa networks: A literature review. *IEEE Communications Surveys & Tutorials*, *23*(3), 1721-1758.
7. Pires, L., & Martins, J. (2024). Experimental Investigation of Spreading Factor, Payload Length and Collision Effects in LoRaWAN Radio Interface.
8. Pires, L., & Martins, J. (2024). Experimental Investigation of Spreading Factor, Payload Length and Collision Effects in LoRaWAN Radio Interface.
9. Bathre, M., & Das, P. K. (2023). Smart dual battery management system for expanding lifespan of wireless sensor node. *International Journal of Communication Systems*, *36*(3), e5389.
10. Kuaban, G. S., Gelenbe, E., Czachórski, T., Czekalski, P., & Tangka, J. K. (2023). Modelling of the energy depletion process and battery

depletion attacks for battery-powered internet of things (iot) devices. *Sensors*, *23*(13), 6183.

11. Kuaban, G. S., Czachórski, T., Gelenbe, E., Sharma, S., Singh, P., Pecka, P., ... & Czekalski, P. (2024). Energy performance of self-powered green iot nodes. *Frontiers in Energy Research*, *12*, 1399371.
12. Da Cruz, M. A., Abbade, L. R., Lorenz, P., Mafra, S. B., & Rodrigues, J. J. (2022). Detecting compromised IoT devices through XGBoost. *IEEE Transactions on Intelligent Transportation Systems*, *24*(12), 15392-15399.
13. Ganesh Raja, M., & Jeyalaksshmi, S. (2024). Self-Configuration and Self-Healing Framework Using Extreme Gradient Boosting (XGBoost) Classifier for IoT-WSN. *Journal of Interconnection Networks*, *24*(03), 2350022.