A Novel Dual Prediction Scheme for Data Communication Reduction in IoT-Based Monitoring Systems

Ahmed Fathalla[1], Ahmad Salah[2], Mohamed Ali Mohamed[3], Nur Indah Lestari[4], and Mahmoud Bekhit[1,5](✉)

[1] Department of Mathematics, Faculty of Science, Suez Canal University, Ismailia 41522, Egypt
{fathalla_sci,mahmoud_bakhit}@science.suez.edu.eg

[2] Faculty of Computers and Informatics, Zagaizg University, Zagazig 44519, Egypt
ahmad@zu.edu.eg

[3] 2P Perfect Presentation, Riyadh 105523, Saudi Arabia

[4] 2 Messiter Street, Campsie, NSW 2194, Australia

[5] School of Electrical and Data Engineering, University of Technology Sydney, Sydney 2007, Australia

Abstract. Internet of things (IoT) based monitoring systems became commonplace. These systems are built upon a large number of devices and sensors. The data collection task of a large number of sensors and devices in an IoT system includes a massive number of data communications. The more the number of devices, the critical is the network bottleneck. In this context, the dual prediction scheme was proposed as a solution for mitigating the large size of communication volumes. The dual prediction scheme consists of a model for predicting future measurements based on historical data. This model is duplicated on both sides, the edge side (i.e., sensor) and the data collection device (i.e., cluster head). The literature includes several works which proposed many dual prediction schemes based on several techniques such as filters and moving average. The literature does not include utilizing the ensemble learning models. This motivates this work to investigate the gradient boosting regression model's performance compared to the existing solutions. The proposed and state-of-the-art models are evaluated on a realistic dataset. The obtained results show that the proposed model outperforms the existing dual prediction schemes in terms of communication reduction.

Keywords: Dual prediction scheme · Gradient boosting · IoT · Monitoring system · Regression

1 Introduction

The Internet of Things (IoT) is a term used to enable the interconnection and communication of various network-enabled devices (e.g., things) that generate

W. Hussain and M. A. Jan (Eds.): IoTaaS 2021, LNICST 421, pp. 208–220, 2022.
https://doi.org/10.1007/978-3-030-95987-6_15

and share data about their surroundings through a network [10,24]. The IoT network is frequently comprised of many sensing devices (*i.e.,* nodes) that are deployed and dispersed across a vast geographical region [13].

Wireless Sensor Networks (WSNs) enable the creation of a diverse variety of possible IoT applications, such as environmental monitoring (e.g., smoke, humidity, temperature, and gas). Sensor devices in the IoT networks capture massive amounts of real-world observation and measurement data. Continuous data transfer between end devices (nodes) results in a massive communication overhead over networks. As a result, the overall energy consumption of IoT applications increases significantly. Reduced data transmission between IoT devices decreases significantly energy consumption and extends the network's lifetime, particularly in battery-powered nodes/networks [4]. Additionally, when identifying a target of interest inside a monitoring surveillance region, it is important to reduce the number of transmissions between sensors and the Fusion Center (FC) in order to maximize network bandwidth efficiency [5]. In such techniques, the FC makes a global determination about the target based on the sensor-transmitted local determination of the target's presence/absence [4].

Data transmission is a significant source of high communication procedures and energy consumption in IoT networks. Additionally, data transmission between end devices utilizes energy than data sensing [12,14,15,22]. In general, network longevity and dependability are critical criteria for many IoT monitoring application scenarios and ensuring that the cloud's resources are utilized to their full potential [19].

While these techniques can considerably reduce transmissions, they suffer from performance degradation to increase the accuracy, necessitating the constant updating of prediction models to include fine-grained changes [22]. Dual prediction techniques [3,23,29] have been developed to overcome this issue. In dual prediction schemes, the prediction model are deployed at the end point nodes. That is, the same prediction model with the same parameters are running on the sensing device and the other device (*e.g.,* cluster head) so that both devices are predicting the same values using historical data of previous data obtained from sensor nodes. Sensor nodes choose whether or not to broadcast their observations to a central server based on the discrepancy between the predicted and real values (e.g., a gateway or a base station).

Dual prediction reduces the number of data transfers between the sensor and the sink nodes to the possible minor level, requiring sensor nodes to communicate just a subset of their felt values without compromising the accuracy of the actual measurement results [8].

Several present research focuses on dual prediction models as proposed in [6,25,26]. Optimizing the communication route between sensors is critical for IoT networks to consume fewer energy [26]. To this aim, IoT networks are shifting their focus to creating local techniques for data transmission reduction and delay avoidance. Numerous publications, such as [4,6,22], used data reduction methods to data gathered by sensor nodes in order to reduce data transmission in sensor networks [25]. Data reduction is frequently accomplished by using predictive models that attempt to anticipate the present measured values of sensor nodes based on the data they have already provided to a Base Station (BS) or a Cluster Head (CH). In this scenario, measurements that may be anticipated are omitted based on specific criteria such as accuracy, resulting in less data transfer between nodes and a central server.

The same scenario applies between sensor nodes and CH; in principle, it is determined by the network topology or its architecture. Interested readers may study the influence of data prediction systems on the decrease of WSN transmissions by consulting Dias et al. [8]. Kalman filters are used by Jain et al. [20] to build a dual prediction method that anticipates future sensor data. As an alternative to pre-existing information about sensor data approaches, Kalman filters depend on a priori assumptions about statistical data characteristics (e.g., data distribution).

To the best of the authors' knowledge, evaluation of ensemble learning models and their performance are not investigated yet. In this context, we proposed the first ensemble learning model to be utilized in dual prediction schemes for communication reduction. We framed the problem of sensors measurement prediction as a time series problem. The historical measurements are used to train the proposed model, and then the proposed gradient boost model is used to predict the future sensors' measurements. The main contributions of this paper can be summarized as follows.

- To our knowledge, we proposed the first dual prediction scheme based on an ensemble learning model.
- We evaluated the performance of the proposed scheme on a real-life dataset. The results outlined that the proposed scheme outperformed the existing methods.

The rest of this paper is organized as follows. Section 2 discusses the related work of dual prediction schemes in the field of IoT. In Sect. 3, the proposed system is exposed. The evaluation and expire mental results of the proposed model are discussed in Sect. 4. Finally, the paper is concluded in Sect. 5.

2 Related Work

In this section, we discuss the existing methods for communication reduction in IoT systems. These utilized methods can be classified based on the prediction approach, such as adaptive filter and deep learning. Besides, the data reduction task can be achieved by combining an additional task, e.g., data compression.

An architecture that combined both edge computing and online learning was proposed to predict the future data for IoT search accurately. This architecture was used to reduce the communication time between the layers of IoT search architecture. Besides, minimize the continuous data queries and the upload operations done by the massively distributed sensors. To selectively report genuine data, edge computing was done at the edge sensors to determine the overall transmission value of all data in the reporting cycle. The cloud layer was primarily in charge of collecting the characteristics of reported data and adapting the related prediction model in real-time [28].

In [16], a cost-aware dual prediction scheme model (CA-DPS) was presented to cut down on data transmission between IoT sensor nodes and the fusion center. The suggested CA-DPS technique selects the strategy that delivers the lowest projected transmission cost from various options within a certain prediction horizon. The future transmission cost was calculated by bootstrapping the model residuals associated with each approach. The suggested technique yields a substantial decrease in the communication required for a given error restriction, according to simulation findings using both synthetic and actual measurement data.

IoT applications consume a high amount of energy because huge data are collected, measured, and transmitted. To reduce the giant consumption of energy and extra load on communication, an adaptive method for data reduction (AM-DR) was presented. The AM-DR approach used a convex combination of two decoupled Least-Mean-Square (LMS) windowed filters with different sizes to predict the next measured values at the source and CHs. Hence, sensor nodes only have to transmit their instant sensed values that differ significantly from the predicted values based on a pre-defined threshold [12].

As machine learning methods are successful utilized in different problem [1,2,9,11,17,18], the authors of [21] employed both approaches as part of a two-tier data reduction architecture. The DP method was used to minimize traffic between CHs and sink nodes, whereas the DC scheme was used to reduce traffic between cluster nodes and CHs. NNs, LSTMs, and OSSLMS machine learning algorithms were used for the DP approach. Many DC techniques were used, such as PCA, T-SVD, and NMF. For both DP and DC schemes, efficient methods were constructed and compared in transmission reduction and accuracy. The ultimate goal of this work is to conserve energy and bandwidth. Dual Prediction (DP) and Data Compression (DC) techniques are employed to reduce the transmission burden across WSNs.

A self-managing WSN architecture was proposed to aid sensor network optimization at the application layer, combining cloud computing, data analysis, and sensor networks. Additional processing should be avoided since sensor nodes have limited resources. Using the Reinforcement Learning (RL) technique QLearning, in reaction to changes in the environment, a new approach was devised to alter the sample intervals of the sensor nodes. It allowed for the optimization of WSNs by calculating the prediction methods (single or dual) and cloud service requirements for the applications. Business models might now focus on sensor data due to the suggested self-managing architecture [7].

For WSNs, a two-stage dual prediction data reduction method was developed. The initial stage was data reduction, which attempted to reduce the number of transfers between the sensor and sink nodes. If the gathered data is redundant, predictable, or incorrect, it will be deleted for transmission. To ensure data dependability, incorrect data at the sensor nodes is deleted and replaced with predicted values at the sink node. The Kalman Filter prediction algorithm was used at the sink node to forecast non-transmitted data from end nodes in the second stage [27].

This discussion outlined that there are several methods proposed for communication reduction using dual prediction schemes. While there are several machine learning models utilized in dual prediction schemes, but to our knowledge, none of these methods utilized the ensemble learning predictive models.

3 The Proposed System

In a Dual Prediction Scheme (DPS), sensing devices/nodes send environmental monitoring data collected to the Gateways (GWs) during an initiation phase in order to pick a prediction model [22]. The GW and sensor nodes may pick their own prediction models; on the other hand, sensing nodes may choose their own prediction method and send the predictive model's parameters values to the GW; or, alternatively, the GW may grant the decision to select the prediction method for the group of sensors. Making a decision regarding which prediction model to use can be made at run-time, and it is not essential to commit to a strategy for the whole lifespan of the WSN.

Sensor nodes can take advantage of their closeness to the data source after selecting (and distributing) the prediction models. When it comes to transmitting actual measurements, they can only do so if their forecasts are incorrect. Sensor nodes and GW may start a new initialization phase and pick new prediction models if the accuracy of predictions is compromised from time to time.

The proposed dual prediction scheme is composed of successive phases where phases' transitions are based on a predefined parameters which represents the number of miss-predicted observations. By the end of each phase, the model parameters are updated to fit final data pattern to start a new phase. Then, the updated model parameters are sent to the sensor node to resume the prediction task.

Before any prediction model is adopted, the startup step guarantees that the GW has comprehensive environmental knowledge. As a result, following this phase, the GW may use the same prediction models as the sensor nodes without transmitting anything new. The sensor nodes' and GW's activities are depicted in Fig. 1. Different prediction methods can be selected frequently depending on the information that is concurrently accessible to sensor nodes and the GW. As a disadvantage, the diversity of prediction models is limited by the sensor nodes' memory and processing capacity constraints. The Least Mean Squares (LMS) technique has produced accurate predictions in simulations in which sensor nodes and GW created their prediction models separately [22,25]. For instance, in one situation, just 10% of the data would be required to monitor room temperature correctly [22].

In the proposed dual prediction scheme, we proposed using a gradient boosting on decision trees (i.e., Catboost[1]) as the predictive model. The

[1] https://catboost.ai/.

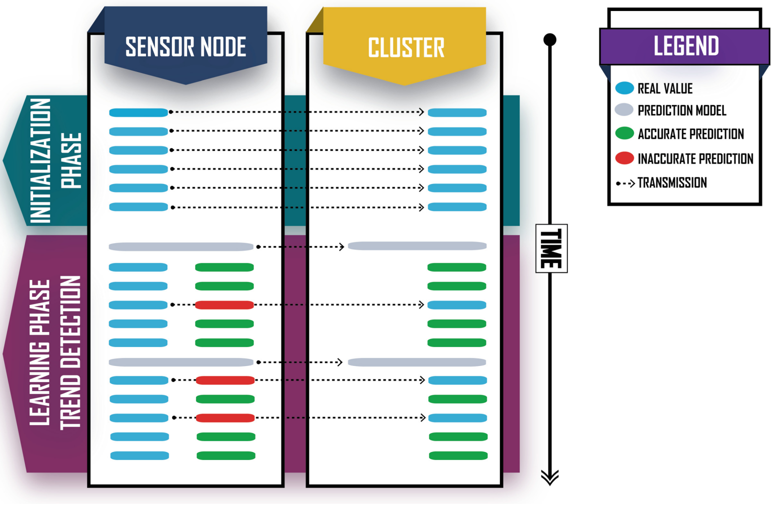


Fig. 1. The proposed dual prediction scheme.

ensemble learning model trains on the historical sensor measurements, and then the trained model is used to predict future measurements. We proposed training the catboost system on lagged data. The lag value in the proposed model is tunable. Thus, the best value can be determined based on the nature of the data patterns. Besides, the update mechanism of the proposed dual prediction scheme depends on the number of predicted measurements. In other words, the user of the proposed scheme can decide when the predictive model is updated/retrained after a tunable number of mispredicted values.

4 Results and Discussion

4.1 Dataset

The dataset is publicly available online[2]. The data are collected from three sensing devices. The collected data are measurements for carbon monoxide (CO), humidity, liquid petroleum gas (LPG), smoke, and temperature. We examined the proposed method performance using 10,000 observations of the smoke values as representative for the other features. In other words, the data transmission reduction rates for the smoke values should be the same for the other features.

4.2 Setup

All the experiments framework are developed in the Python Programming language. We compared the proposed system against two well-known state-of-the-art methods, namely LMS [22], and AM-DR [12]. We used a lag value equal to

[2] https://www.kaggle.com/garystafford/environmental-sensor-data-132k.

three for all methods of comparison. Besides, we set the filter weights for the LMS as the default reported values, and we used the weight of 4 and 8 for the AM-DR method.

Experiments are conducted using three IoT sensing devices to evaluate the proposed model performance when data from different devices are used. The proposed approach and the comparative methods of comparison are evaluated in terms of transmission reduction percentage of the environmental data when different error-thresholds are used, the number of experiment's phases, and the number of miss-predictions. Data reduction percentage metric is defined in the base of Eq. 1.

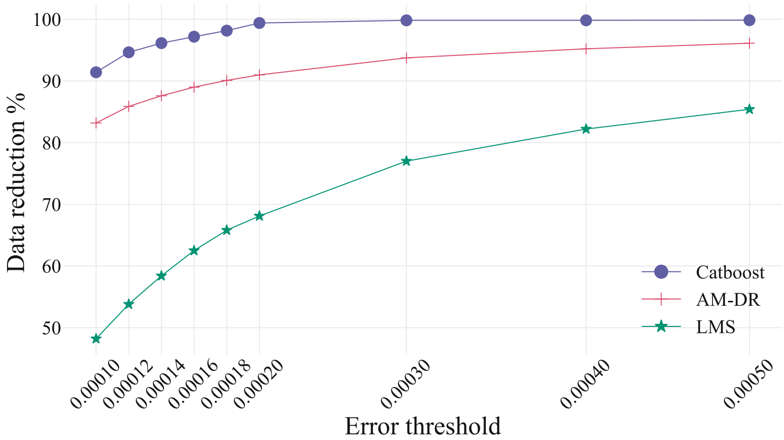
$$data\_reduction\_ratio = \frac{NT}{S} \times 100 \tag{1}$$

where NT denotes the not transmitted observations (miss-predictions) count, and S is the total number of sensed observation.

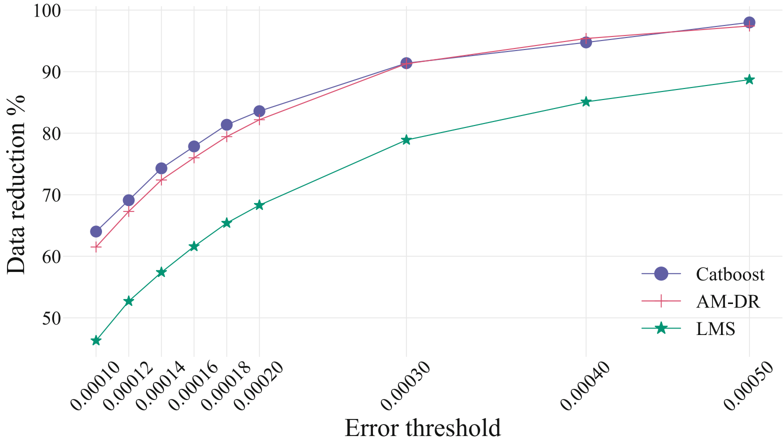
4.3 Experimental Results

The first experiment examines the data reduction percentages when various error-threshold values are employed. For that purpose, we tested a wide range of error-threshold for three different sensors (i.e., sensing devices), as shown in Fig. 2. The Figure depicts the data reduction percentages for forecasting smoking values. In this experiment, the model is trained using $1,000$ observations and is tested using $10,000$ observations. Obviously, the higher the tolerance (i.e., threshold value), the higher the obtained data reduction rates. The proposed predictive model achieved the highest reduction rates compared to the other comparison methods for the three devices. Figure 2 clarifies that the proposed scheme has a significant performance gap compared to the LMS methods. On the other hand, the performance gap between the proposed and AM-DR methods varies for the different utilized devices. However, the overall performance of the proposed method outperformed the AM-DR method.

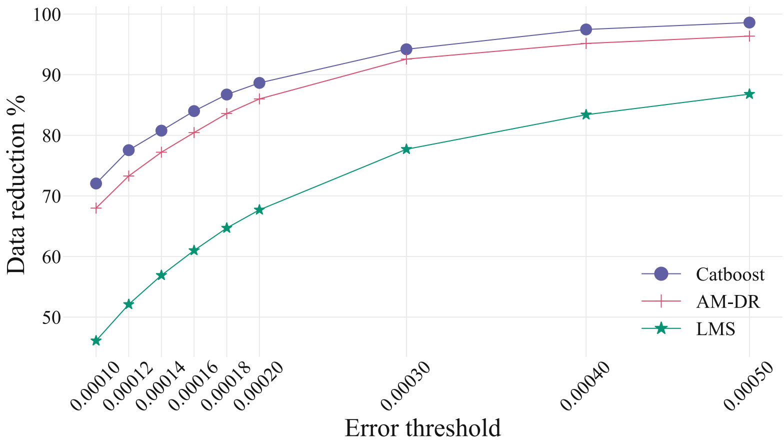
The second part of the results is about the performance of the proposed predictive model under different parameter values. We studied the effect of selecting a different number of mispredicted measurements on the data reduction and the number of model updates (i.e., phases) in Fig. 3. In Fig. 3(a), when the predictive model is updated after one mispredicted value, the data reduction was about 94.4%. Similarly, when the experiment was run with updating the predictive model after each five mispredicted values, the data reduction rate increased to almost 95%. Then, the data reduction rate decreased kept fluctuating as the number of mispredictions to update the model parameter increased. Thus, there is no clear relationship to decide the optimal number of mispredicted values to update the model. The user should examine this value to pick the optimal value based on the nature of the data at hand. In Fig. 3(b), the relationship between the number of times the model needed to be updated (i.e., phases), the y-axis, in correspondence to the number of mispredicted values needed to update the model, the x-axis. It is clear from Fig. 3 that the higher is the allowed number



(a) Device 1

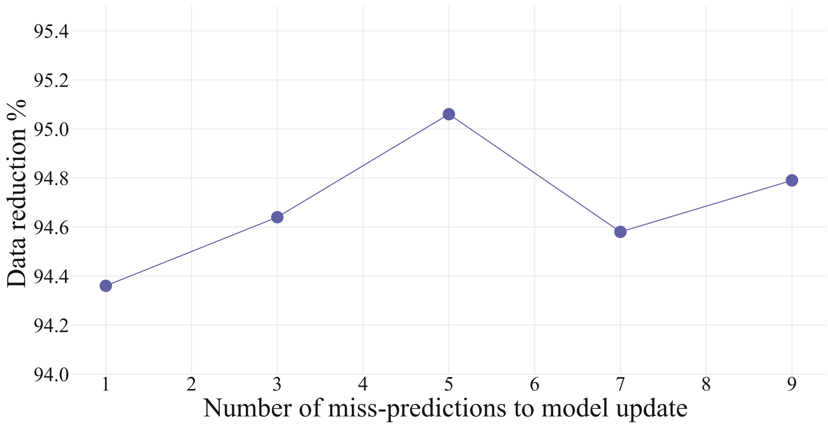


(b) Device 2

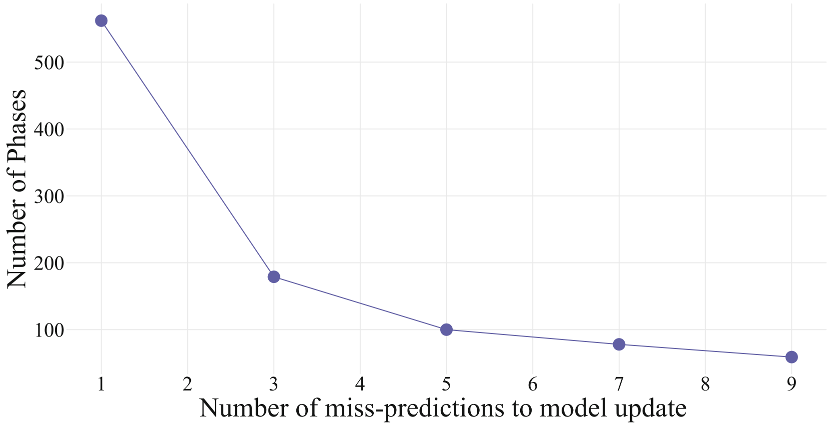


(c) Device 3

Fig. 2. Comparison of the proposed method, AR-DR, and LMS filters on data reduction rates.



(a) Number of mispredictions for updating the model vs. the data reduction rate



(b) Number of misprediction for updating the model vs. the number of phases

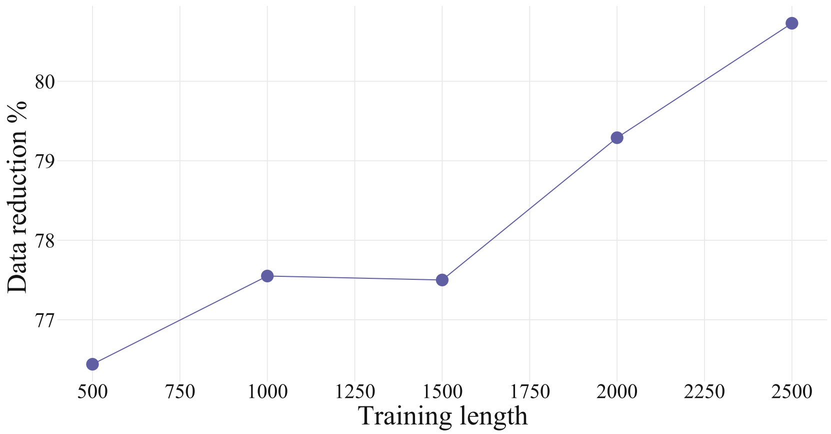
Fig. 3. The effect of different number of misprediction for updating the model on the performance.

of mispredicted values, the less is the number of required times to update the model.

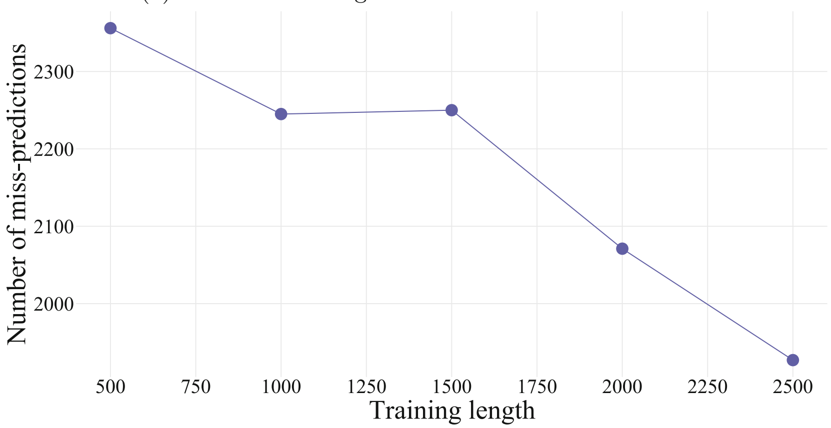
In addition, we validated the accuracy of the model under different volumes of the training dataset in Fig. 4. In Fig. 4(a), the data reduction increases as the training data increases; This is linked to the model accuracy. The more is the data; the better is the model accuracy. Similarly, in Fig. 4(b), the number of mispredicted readings decreases as the training data increases.

4.4 Limitations

The proposed scheme has certain restrictions and underlying assumptions that must be considered. As a starting point, we present the scheme's assumptions.



(a) Different training sizes vs. data reduction rates.



(b) Different training sizes vs. number of mispredictions for model updating.

Fig. 4. The effect of training size on the model accuracy.

Fog and edge computing architectures are suggested as WSN devices have limited processing power and energy. For that purpose, We take into account an IoT system in which data is exchanged between a cluster head and a sensing node. Generally, the proposed scheme considers similar situations where communication takes place between two end nodes communicating together with such comparable fog or edge computing architectures. One of the main limitations is that we suggest that the cluster head endpoint has an active and direct connection to a powerful and battery-powered device (i.e., fog device) to process the data and develop the required machine learning model. Meanwhile, the other endpoint has a the ability to process data, or is connected to an external edge device (e.g., Raspberry Pi).

The second limitation is that with the assumptions mentioned above are met, the training and updating procedures of the model might resume for few seconds.

Thus, that may result in a few or ten/hundred missed observations for specific time windows. Another issue is that the machine learning model must be sent from the fog device to the edge device, which takes kilobytes of data to simulate the energy use during this procedure.

5 Conclusion

The communication volumes in IoT-based monitoring systems are massive. One possible solution for this bottleneck in these monitoring systems is the dual prediction scheme. In this scheme, the sensors and CH predict the sensor's future measurements using the same predictive model. Only miss-predicted measurements are transferred from the sensing device to the CH if the difference between the real and predicted measurements is higher than a predefined threshold. In this vein, we proposed an ensemble learning model (i.e., gradient boost) as the predictive model in this scheme. The proposed predictive model performance is evaluated against state-of-the-art methods on a real-life environmental data of various attributes. The proposed model achieved a communication reduction reached 99%, which outperformed the existing methods.

References

1. Alameer, Z., Fathalla, A., Li, K., Ye, H., Jianhua, Z.: Multistep-ahead forecasting of coal prices using a hybrid deep learning model. Resour. Policy **65**, 101588 (2020)
2. Ali, A., Fathalla, A., Salah, A., Bekhit, M., Eldesouky, E.: Marine data prediction: an evaluation of machine learning, deep learning, and statistical predictive models. Comput. Intell. Neurosci. **2021** (2021)
3. Almalki, F.A., Ben Othman, S., Almalki, F.A., Sakli, H.: EERP-DPM: energy efficient routing protocol using dual prediction model for healthcare using IoT. J. Healthcare Eng. **2021** (2021)
4. Babcock, B., Olston, C.: Distributed top-K monitoring. In: Proceedings of the 2003 ACM SIGMOD International Conference on Management of Data, pp. 28–39 (2003)
5. Ciuonzo, D., Rossi, P.S.: Quantizer design for generalized locally optimum detectors in wireless sensor networks. IEEE Wireless Commun. Lett. **7**(2), 162–165 (2017)
6. Deshpande, A., Guestrin, C., Madden, S.R., Hellerstein, J.M., Hong, W.: Model-driven data acquisition in sensor networks. In: Proceedings of the Thirtieth International Conference on Very Large Data Bases, vol. 30, pp. 588–599 (2004)
7. Dias, G.M., Bellalta, B., Oechsner, S.: Using data prediction techniques to reduce data transmissions in the IoT. In: 2016 IEEE 3rd World Forum on Internet of Things (WF-IoT), pp. 331–335. IEEE (2016)
8. Dias, G.M., Bellalta, B., Oechsner, S.: The impact of dual prediction schemes on the reduction of the number of transmissions in sensor networks. Comput. Commun. **112**, 58–72 (2017)
9. Eldesouky, E., Bekhit, M., Fathalla, A., Salah, A., Ali, A.: A robust UWSN handover prediction system using ensemble learning. Sensors **21**(17), 5777 (2021)

10. Fathalla, A., Li, K., Salah, A., Mohamed, M.F.: An LSTM-based distributed scheme for data transmission reduction of IoT systems. Neurocomputing **485**, 166–180 (2021)
11. Fathalla, A., Salah, A., Li, K., Li, K., Francesco, P.: Deep end-to-end learning for price prediction of second-hand items. Knowl. Inf. Syst. **62**(12), 4541–4568 (2020). https://doi.org/10.1007/s10115-020-01495-8
12. Fathy, Y., Barnaghi, P., Tafazolli, R.: An adaptive method for data reduction in the internet of things. In: 2018 IEEE 4th World Forum on Internet of things (WF-IoT), pp. 729–735. IEEE (2018)
13. Fathy, Y., Barnaghi, P., Tafazolli, R.: Large-scale indexing, discovery, and ranking for the internet of things (IoT). ACM Comput. Surv. (CSUR) **51**(2), 1–53 (2018)
14. Gamal, M., Abolhasan, M., Lipman, J., Liu, R.P., Ni, W.: Multi objective resource optimisation for network function virtualisation requests. In: 2018 26th International Conference on Systems Engineering (ICSEng), pp. 1–7. IEEE (2018)
15. Gamal, M., Morsy, E., Fathy, A.: Multi-objective transmitters placement problem in wireless networks. In: Proceedings of the Sixth International Symposium on Information and Communication Technology, pp. 156–162 (2015)
16. Håkansson, V.W., Venkategowda, N.K., Kraemer, F.A., Werner, S.: Cost-aware dual prediction scheme for reducing transmissions at IoT sensor nodes. In: 2019 27th European Signal Processing Conference (EUSIPCO), pp. 1–5. IEEE (2019)
17. Hussain, W., Hussain, F.K., Saberi, M., Hussain, O.K., Chang, E.: Comparing time series with machine learning-based prediction approaches for violation management in cloud SLAS. Futur. Gener. Comput. Syst. **89**, 464–477 (2018)
18. Hussain, W., Merigó, J.M., Raza, M.R.: Predictive intelligence using ANFIS-induced OWAWA for complex stock market prediction. Int. J. Intell. Syst. (2021)
19. Hussain, W., Merigo, J.M., Gao, H., Alkalbani, A.M., Rabhi, F.A.: Integrated AHP-IOWA, POWA framework for ideal cloud provider selection and optimum resource management. IEEE Trans. Serv. Comput. (2021)
20. Jain, A., Chang, E.Y., Wang, Y.F.: Adaptive stream resource management using Kalman filters. In: Proceedings of the 2004 ACM SIGMOD International Conference on Management of Data, pp. 11–22 (2004)
21. Jarwan, A., Sabbah, A., Ibnkahla, M.: Data transmission reduction schemes in WSNs for efficient IoT systems. IEEE J. Sel. Areas Commun. **37**(6), 1307–1324 (2019)
22. Santini, S., Romer, K.: An adaptive strategy for quality-based data reduction in wireless sensor networks. In: Proceedings of the 3rd International Conference on Networked Sensing Systems (INSS 2006), pp. 29–36. TRF, Chicago (2006)
23. Shu, T., Chen, J., Bhargava, V.K., de Silva, C.W.: An energy-efficient dual prediction scheme using LMS filter and LSTM in wireless sensor networks for environment monitoring. IEEE Internet Things J. **6**(4), 6736–6747 (2019)
24. Sohaib, O., Lu, H., Hussain, W.: Internet of things (IoT) in e-commerce: for people with disabilities. In: 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA), pp. 419–423. IEEE (2017)
25. Stojkoska, B., Solev, D., Davcev, D.: Data prediction in WSN using variable step size LMS algorithm. In: Proceedings of the 5th International Conference on Sensor Technologies and Applications, pp. 191–196. Citeseer (2011)
26. Stojkoska, B.L.R., Trivodaliev, K.V.: A review of internet of things for smart home: challenges and solutions. J. Clean. Prod. **140**, 1454–1464 (2017)
27. Wang, H., Yemeni, Z., Ismael, W.M., Hawbani, A., Alsamhi, S.H.: A reliable and energy efficient dual prediction data reduction approach to WSNs based on Kalman filter. IET Commun. **15**, 2285–2299 (2021)

28. Wu, F., Chen, Y., Chen, X., Fan, W., Liu, Y.: An adaptive dual prediction scheme based on edge intelligence. IEEE Internet Things J. **7**(10), 9481–9493 (2020)
29. Yu, T., Wang, X., Shami, A.: A novel fog computing enabled temporal data reduction scheme in IoT systems. In: GLOBECOM 2017–2017 IEEE Global Communications Conference, pp. 1–5. IEEE (2017)