Comparison Analysis of Sensor Data Prediction-Based IoT using MLP in Digital Twin

Adinda Riztia Putri, Made Adi Paramartha Putra, Jae Min Lee, and Dong-Seong Kim
Department of IT Convergence Engineering, Kumoh National Institute of Technology, Gumi, South Korea
(adindariztia, mdparamartha95, lmjpaul, dskim)@kumoh.ac.kr

***Abstract*—Data prediction-based mechanism in predicting the sensor output is gaining more attention due to their ability in limiting or lowering the data transmission rate between the sensor nodes and the edge node. Digital twins are also using data prediction methods to determine the sensor output for its logical object counterparts. In this study, the basic configuration of how many data points are to be predicted solely on the prediction model itself is investigated. Our result shows that the proposed model can accurately predict 10 data points continuously without degrading its performance.**

***Index Terms*—Data recovery, Deep Learning, IoT, WSN.**

I. Introduction

Data prediction in Wireless Sensor Networks (WSN) is gaining more attention as the attention on energy efficiency issue gaining more spotlight. The sensor's nature of continuously transducing physical signal to electrical signal as data, to then transmit the processed signal to other nodes makes the sensor consume the battery energy faster. This is because data transmission is one of the most energy-exhaustive operations of the sensor. Further reducing or limiting the data transmission operation on the sensor node is proven to be one of the most effective approaches to slow the drainage rate of the sensor's battery [1].

Later, the data prediction model is adopted as one of the mechanisms in predicting the sensor's output implemented in a virtual sensor, a digital twin representative of a sensor. In digital twin, the real physical sensor is replicated as a local object where it performs the physical object counterparts' function. In short, the data output mechanism of the sensor is performed by the data prediction model of the deep learning model. By implementing this strategy, the data transmission of the model hence can be limited and thus maintaining the sensor energy consumption.

However, a sensor is sensitive to Spatio-temporal change in its deployed environment. Therefore, trusting sensor output that is solely based on the data prediction model for the long run is not recommended. Eventually, the real sensor data is needed to adjust the model output and keep the sensing mechanism between the logical and physical object of the digital twin in track. Bidirectional data communication between the logical object and the physical object is thus needed to exchange information about the current environmental reading from the physical object versus the logical object.

Several studies have been performed in the past to perfect the data prediction model for sensor data prediction. LSTM was mainly used as the most used approach in predicting the sensor output due to its infamous ability to see patterns in past data and accurately forecast future data [2] [3] [4]. However, LSTM is also infamous for its resource-consuming architecture, thus adding computational overhead for the machine. Unfortunately, this situation is not suitable for the device-constraint environment such as in IoT. Many studies have performed a cloud-based approach to split the heavy process to the more powerful machine as in the cloud server. However, this approach will also increase the network latency by up to 20 seconds [5].

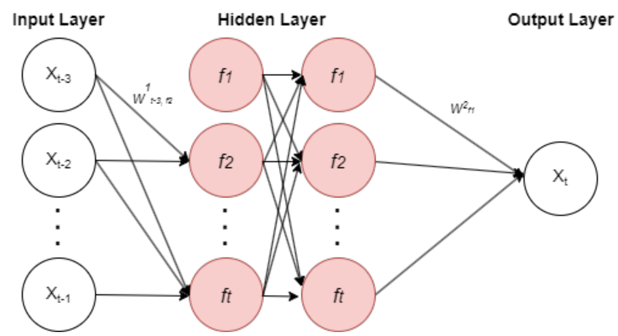

Fig. 1: Multi Layer Perceptron neural network configuration

Multi-Layer Perceptron has gained popularity for its simple architecture yet powerful performance. Due to its simplicity, the processing time is faster compared to LSTM. Several studies have shown that MLP performance in predicting sensor output is reliable. However, it is unclear how much output it can produce to support the digital twin logical object functions before it shows decreased performance trends. In this study, we want to highlight MLP performance in predicting the sensor output streams and how the number of predictions affects the model performance.

The rest of this study is organized as follows. Section II describes the proposed method underlying the data prediction model in this study. Section III discussed the performance analysis comparison between the obtained result in performing the data prediction tasks for digital twins. Finally, Section IV concludes the result of this study.

II. Proposed System

The proposed method for this study is based on the MLP architecture depicted in Figure 1. In general, or MLP model

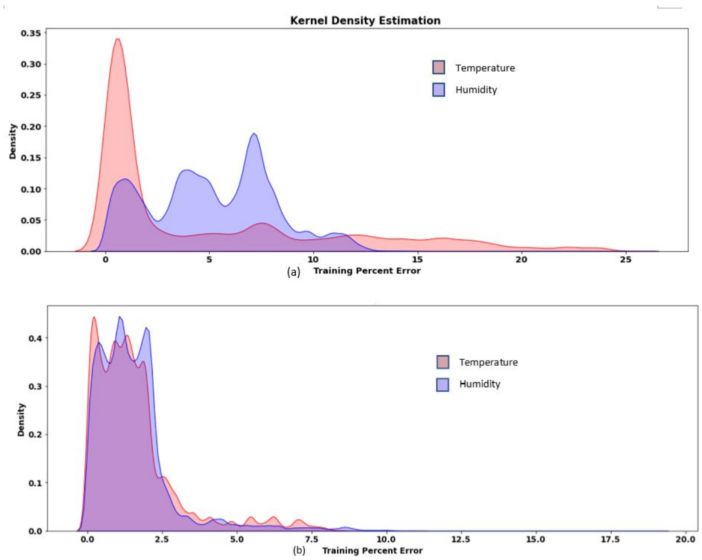

Fig. 2: Distribution of the training percentage error in KDE

has initial layer with the size of $\{X_{t-3},...,X_{t-1}\}$ number of neurons. These initial layers take the input of t number sensor data which also acted as the timesteps data. After processing the data in the first layer, the data is passed through the hidden layer which each has a bias value equals to 1 [6]. The backpropagation mechanism in MLP updates the weight value of the neurons based on the expected output that is visible to the model during the training process. Finally, on the X_t neuron, the output data value of the sensor is predicted.

The DHT-22 sensor dataset is used for the training and evaluation process. The dataset contains 10 days' worth of temperature and humidity measurements that the sensors took. The sensor took 2 data points per second, namely the temperature and the humidity. This dataset is then split in a ratio of 7:3 for training vs testing.

III. Performance Evaluation

This section discusses the performance of the proposed MLP model in predicting the various number of data points continuously. The performance is analyzed based on their respective RMSE, MAE, and MAPE value.

From Table 1, the performance of the proposed method is explained from various data points to see how much a model can properly keep making predictions before the model performance is compromised. As we can see, Table 1 indicates that the more data is being predicted by the model, the more degrading its performance. In Figure 2 a, more data points are being predicted continuously (20 data) compared to Figure 2 b (5 data). The error percentage for the previous configuration is higher compared to the latter one. The density and error rate can be compressed down maximally if the smaller data points configuration is chosen.

For the optimal point, we find that by predicting 10 data points continuously, the model still retains its best performance. This is important as for the digital twin implementation, the more data a model can predict may result in the more energy to be saved. This is because of having a such condition where the logical object counterpart can independently predict the data by itself without the data transmission from the physical object part.

TABLE I: The performance comparison of the MLP model in predicting various continuously datapoints

MLP Output	RMSE	MAE	MAPE
5 datapoints	0.115	0.084	0.063
10 datapoints	0.116	0.08	0.069
15 datapoints	0.134	0.099	0.075
20 datapoints	0.14	0.12	0.081

IV. Conclusion

This study presented a comparison analysis in determining how many data points are to be predicted by a data prediction model based on MLP in a digital twin environment. This is necessary for setting up the initial configuration of how the bidirectional communication between the physical and logical objects in digital twins. Based on the obtained result, predicting 10 data points continuously in a digital twin is more beneficial for the system as it does not sacrifice the model performance and still limits the transmission rate between the sensor node and another node. The best performance obtained by the model is 0.116, 0.08, and 0.069 for RMSE, MAE, and MAPE respectively. For future study, the implementation of the proposed method in a digital twin is to be investigated, along with how the model affected the transmission efficiency.

Acknowledgement

This research work was supported by Priority Research Centers Program through NRF funded by MEST(2018R1A6A1A03024003) and the Grand Information Technology Research Center support program (IITP-2022-2020-0-01612) supervised by the IITP by MSIT, Korea.

References

[1] T. Wang, D. Zhao, S. Cai, W. Jia, and A. Liu, "Bidirectional Prediction-Based Underwater Data Collection Protocol for End-Edge-Cloud Orchestrated System," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 7, pp. 4791–4799, 2020.

[2] A. P. Hermawan, D.-S. Kim, and J.-M. Lee, "Sensor Failure Recovery using Multi Look-back LSTM Algorithm in Industrial Internet of Things," in *2020 25th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, 2020, pp. 1363–1366.

[3] M. A. P. Putra, A. P. Hermawan, D.-S. Kim, and J.-M. Lee, "Energy Efficient-based Sensor Data Prediction using Deep Concatenate MLP," in *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*, 2021, pp. 1–4.

[4] M. A. P. Putra, D.-S. Kim, and J.-M. Lee, "DB-BiLSTM: Euclidean Distance-Based Sensor Data Prediction for IoT Applications," in *2021 International Conference on Information and Communication Technology Convergence (ICTC)*, 2021, pp. 814–817.

[5] R. Saha, A. Chakraborty, S. Misra, S. K. Das, and C. Chatterjee, "DLSense: Distributed Learning-Based Smart Virtual Sensing for Precision Agriculture," *IEEE Sensors Journal*, vol. 21, no. 16, pp. 17 556–17 563, 2021.

[6] J. M. N. Abad and A. Soleimani, "Novel Feature Selection Algorithm for Thermal Prediction Model," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 26, no. 10, pp. 1831–1844, 2018.