POAC'23
Glasgow, UK

Proceedings of the 27th International Conference on
Port and Ocean Engineering under Arctic Conditions
12~14, Glasgow, UK

Estimating Missing Hull Strain Gauge Data of the ARAON Using Artificial Intelligence

Eunjin Oh[1], Jeong Suk Ha[1]
[1] Korea Research Institute of Ships and Ocean
[2] Daejeon, Republic of Korea

ABSTRACT

The hull strain is measured by using a strain gauge located to the inner plate of IBRV Araon hull. This procedure is conducted annually to estimate local ice load. At this time, some values of the sensor are frequently missing. These missing values have made it difficult to determine the local ice load. In this study, a interpolation method is proposed to estimate the missing value of the hull strain sensor for estimation the local ice load during icebreaking. Due to the influence of the sensor value in which the missing value occurred, it is difficult to accurately estimate the magnitude of the ice load with the influence coefficient matrix method. In addition, it is not easy to reveal the correlation between each sensor value. In order to estimate missing values, a study was conducted to estimate the missing values of the sensor by learning the correct sensor values using the artificial neural network deep learning method.

KEY WORDS Ice; Missing data; Hull Strain Gauges; LSTM; Icebreaker.

1. Introduction

The ARAON is a South Korean icebreaking research vessel designed to support polar exploration and scientific research. Hull strain gauges are installed on the ship to monitor stress levels experienced by its structure during operations. However, missing data in the recorded strain gauge data can cause issues in accurate analysis and interpretation. This paper investigates the use of artificial intelligence (AI) techniques to estimate missing data from the Hull strain gauge data of the ARAON.

The maritime industry has long recognized the importance of monitoring the structural integrity of ships to ensure their safety during operations. One of the critical components in maintaining a ship's structural integrity is monitoring the stress experienced by the ship's hull under various operational conditions. Hull strain gauges are widely used to measure strain experienced by a ship's structure due to external forces such as hydrodynamic pressure, ice loads, and wave impacts. The accurate measurement and analysis of strain data are crucial for predicting potential structural failures, optimizing ship design, and ensuring safe and efficient maritime operations.

The ARAON, a South Korean icebreaking research vessel, is designed to support polar exploration and scientific research in the Arctic and Antarctic regions. As a ship operating in harsh environments, the ARAON is equipped with Hull strain gauges to monitor its structural integrity. However, missing data in the recorded strain gauge data can cause inaccuracies in the analysis and interpretation of the strain measurements, which may ultimately compromise the safety and efficiency of the vessel's operations.

With the recent advancements in artificial intelligence (AI) and machine learning, there is a growing interest in harnessing the potential of these technologies to solve various challenges in the maritime industry, including the estimation of missing data in Hull strain gauge records. This paper presents a study on using AI techniques to estimate missing data from the Hull strain gauge data of the ARAON. By utilizing AI-based methods, this research aims to improve the accuracy and reliability of strain gauge data analysis, contributing to safer and more efficient maritime operations.

In the following sections, we provide a background on Hull strain gauges and the challenges associated with missing data. We then describe the methodology used in this study, including data collection, preprocessing, and the AI techniques investigated for estimating the missing data. The results of the study are presented and discussed, followed by a conclusion highlighting the potential of AI-based methods in addressing the missing data problem in strain gauge data analysis.

2. Background

2.1 Hull Strain Gauges

Hull strain gauges are devices that measure the deformation or strain experienced by a ship's structure under various loads. They typically use electrical resistance-based sensors, such as foil or semiconductor strain gauges, to convert the strain experienced by the ship's hull into an electrical signal that can be measured and recorded. The strain gauges are strategically placed on the ship's hull to capture essential information about the stress experienced by the structure during operations.

2.2 Study with Missing Data in Strain Gauge Data

The influence coefficient method is used to calculate the glacial load on ships, and since matrix calculation is performed at this time, when missing data occurs, it can cause calculation performance problems. Missing data in strain gauge records can occur due to a range of factors such as sensor malfunction, gauge corruption, data transmission errors, or data logging problems. The absence of data points can lead to inaccurate strain analysis, degradation in data quality, and increased uncertainty in estimating strain values, which can affect the reliability of fatigue assessments and structural integrity monitoring. Table 1 illustrates the data generation form of missing data at a particular point. It is crucial to mitigate missing data issues in strain gauge records to ensure reliable analysis of strain behavior in ship structures and inform appropriate decisions on the ship's structural condition or operational strategy.

Table 1. Sensor measurement data status (M : Missing data)

Sensor No.	L_A	L_B	L_C	R_A	R_B	R_C

01	0.6348	0.4273	0.4903	0.6700	0.7840	0.5980
02	0.7324	M	0.7948	0.6857	0.7776	0.6468
03	M	0.4828	0.6717	0.6717	0.7912	0.5481
04	0.7364	0.6607	0.7030	0.6594	M	M
05	0.8224	0.6778	0.6864	0.8065	0.7565	M
06	0.8162	0.5701	0.8892	0.7065	M	0.6465
07	0.7564	0.7114	0.9161	0.8226	0.5565	0.5682
08	0.8511	0.7650	0.9108	0.8272	0.7362	0.5510
09	0.8661	0.7277	0.5606	0.3975	0.4257	0.3922
10	0.7927	0.6883	0.7797	M	0.5119	0.4089
11	0.7384	0.8170	0.7602	0.4241	0.6689	0.6121
12	0.7116	0.7883	0.6351	0.3594	0.6429	0.5923
14	0.7413	0.6987	0.7651	0.6895	0.6895	0.8895

3. Methodology

3.1 Data Collection

The ARAON measurement, Korea's first icebreaker research ship, "Real ship test for measuring ice load and ice resistance in ice areas such as Antarctica since the construction of the Arctic Ocean in 2009" (Joe, Choi, 2019). In this study, only the strain data was extracted from the dataset and analyzed. The hull strain gauge data was collected from the ARAON vessel between June and July 2021, and the dataset includes strain sensor values, information on the ship's operational status, as well as GPS data. In order to indirectly estimate the local ice load on the bow of the ship through a solid line test, a strain sensor is attached to the inner plating behind the collision bulkhead to build a measurement system. The strain sensor measures the three axes of strain using a fiber optic sensor and a rosette gauge. As shown in Figure 1, a total of 25 strain gauge sensors are attached to the port and starboard, collecting information at a transmission rate of 50Hz

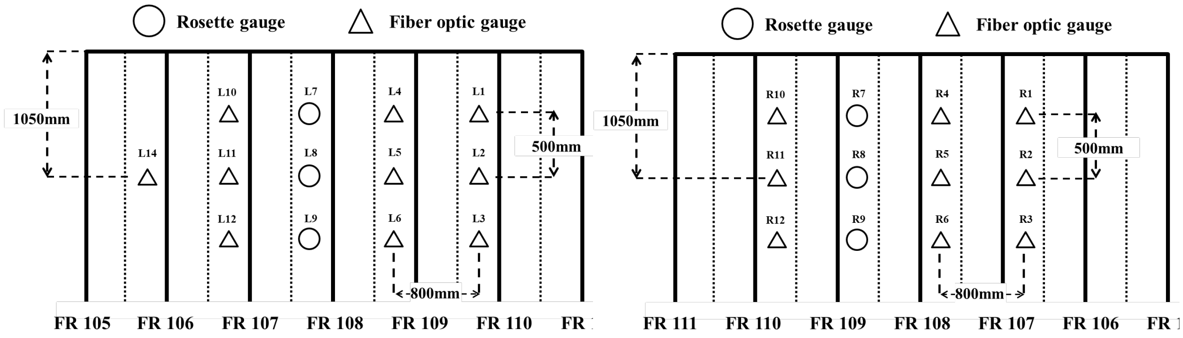
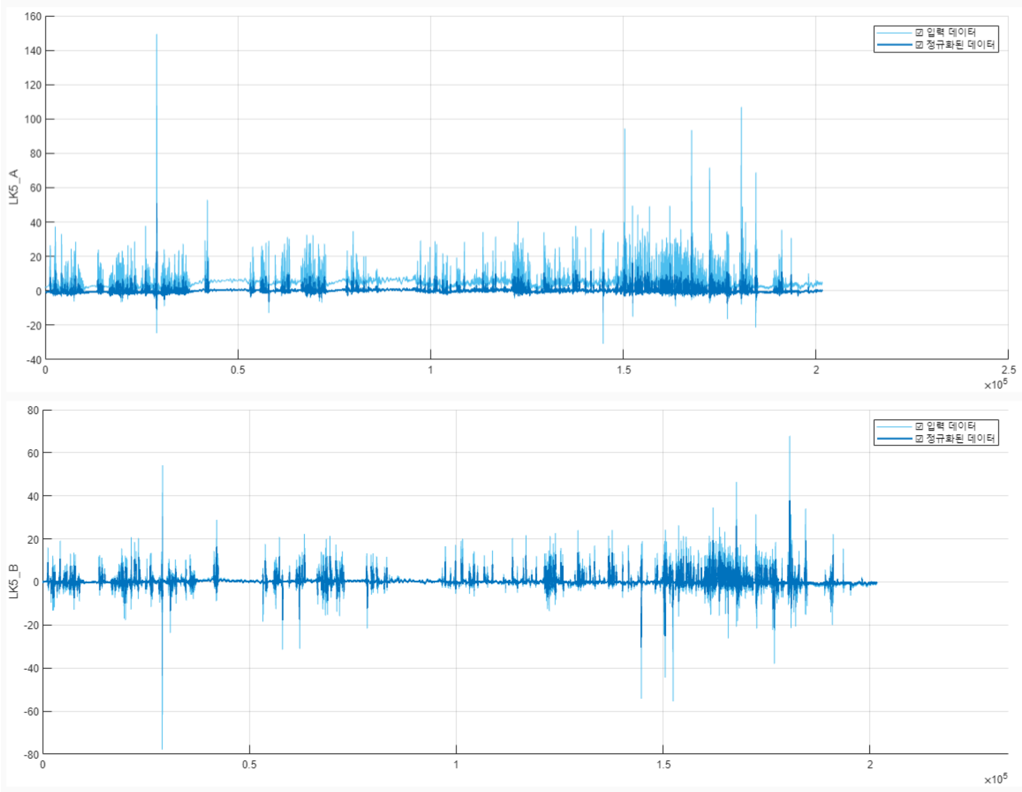


Figure 1. Port side(Left) and Starboard side(Right) senor position

3.2 Preprocessing

The preprocessing stage involved several steps to clean, normalize, and prepare the dataset for training and evaluation of the AI-based methods. These steps included:

1. Data cleaning: The dataset was checked for inconsistencies, such as duplicate entries, incorrect data types, and unrealistic values. Any identified issues were corrected, ensuring the dataset's quality and integrity.
2. Outlier detection and removal: Outliers in the strain gauge readings can negatively impact the performance of AI models. Statistical methods, such as the interquartile range (IQR) method and Z-score, were employed to identify and remove potential outliers in the dataset. In our research, we utilized the Z-score method.
3. Missing data identification: Missing data points in the strain gauge readings were identified and labeled. These missing data points were the target of estimation by the AI-based methods.
4. Feature engineering: Additional features were derived from the raw data to provide more context for the AI models. These features included time-based and derived variables from the operational data
5. Data normalization: To ensure that all features were on a comparable scale and to improve the performance of the AI models, the data were normalized using methods such as Min-Max scaling and Z-score normalization.
6. Data partitioning: The dataset was divided into training, validation, and testing sets. The training set was used to train the AI models, the validation set was used for model selection and hyperparameter tuning, and the testing set was used to evaluate the models' performance in estimating missing data points.
7. 

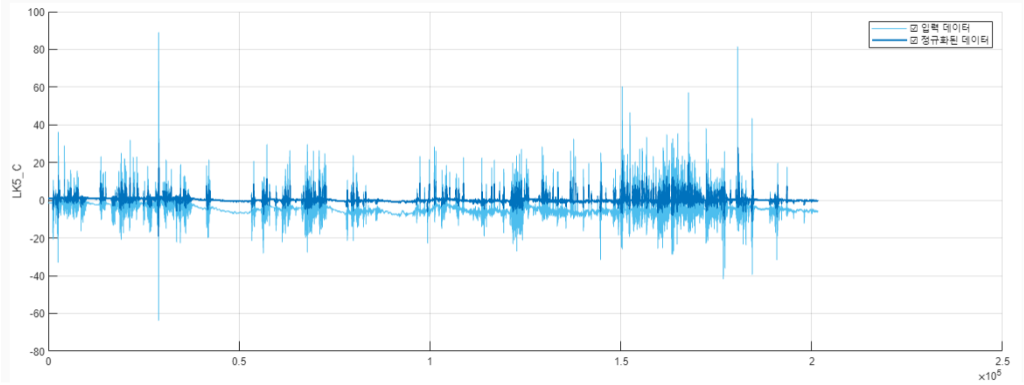


Figure 2. Detect and remove outliers and normalize data

3.3 AI Techniques for Estimating Missing Data

In this study, Three AI techniques were investigated for estimating missing data in the Hull strain gauge data of the ARAON: Linear Regression, K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM) neural networks.

3.3.1 Linear Regression

Linear regression is a statistical method that models the relationship between a dependent variable (in this case, the strain gauge reading) and one or more independent variables (e.g., time, operational data, and other strain gauge readings). The model estimates the missing data by fitting a line or hyperplane that best describes the relationship between the dependent and independent variables.

The general formula for a linear regression model with multiple independent variables is:

$$y = \beta 0 + \beta 1x1 + \beta 2x2 + ... + \beta kxk + \varepsilon \qquad (1)$$

where:

1. y is the dependent variable (strain gauge reading)
2. β0 is the intercept term (the value of y when all independent variables are zero)
3. β1, β2, ..., βk are the coefficients for the independent variables x1, x2, ..., xk (which represent the change in y for a one-unit change in the corresponding independent variable)
4. ε is the error term (which accounts for the difference between the actual and predicted values of y)

In the context of the Hull strain gauge data of the ARAON, the independent variables (x1, x2, ..., xk) might include time, operational data, and other strain gauge readings. The goal of linear regression is to find the optimal values for the coefficients (β1, β2, ..., βk) that minimize the sum of the squared differences between the actual strain gauge readings and the predicted readings. This is typically achieved using the least-squares method.

The linear regression model was built using the following steps:

1. Feature selection: Feature selection: The relevant independent variables were identified through a correlation analysis, which determined the strength and direction

of the relationships between the dependent variable (strain gauge reading) and the potential independent variables.

2. Model fitting: The linear regression model was fitted using the least-squares method, which minimizes the sum of the squared differences between the actual strain gauge readings and the predicted readings. This method finds the optimal weights for each independent variable, allowing the model to make accurate predictions for missing data points.

3. Model validation: The model's performance was assessed using metrics such as mean squared error (MSE), mean absolute error (MAE), and R-squared (R2). The model's performance was also compared with other AI-based methods to determine the most suitable technique for estimating missing data in the strain gauge data.

4. Missing data estimation: The fitted linear regression model was used to estimate the missing data points in the strain gauge data. The model takes the values of the independent variables as inputs and provides an estimate of the corresponding strain gauge reading.

While linear regression is a simple and interpretable method, it has certain limitations, such as assuming a linear relationship between the dependent and independent variables and being sensitive to outliers and multicollinearity. These limitations might impact the model's accuracy and reliability when estimating missing data in the strain gauge data.

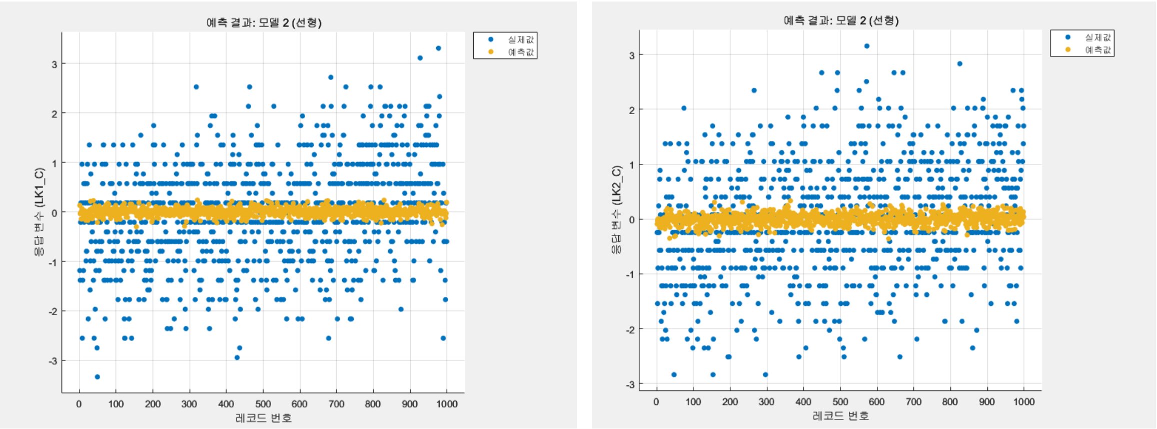


Figure 3. Linear regression model estimation results

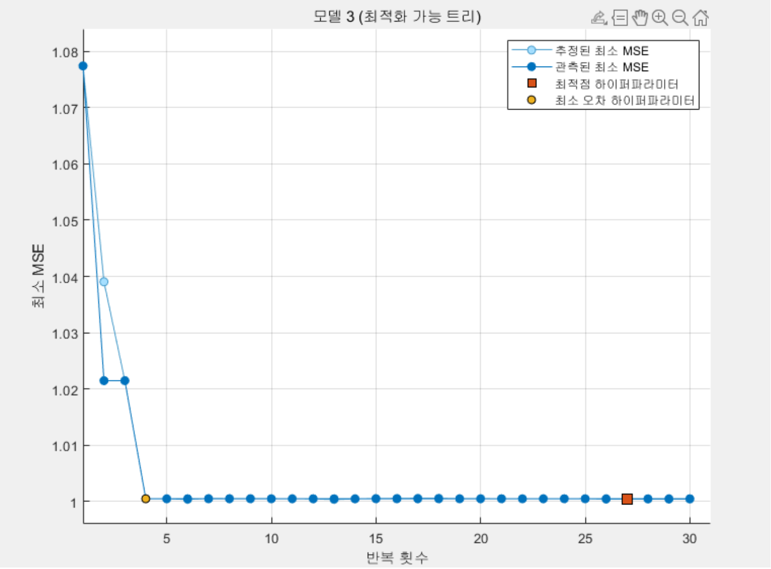


Figure 4. MSE for linear regression model

3.3.2 K-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) is a non-parametric method used for classification and regression tasks. In the context of estimating missing data, KNN is used as a regression technique. The main idea behind KNN is to estimate the missing value based on the values of its K nearest neighbors in the feature space. The missing value is estimated by taking the weighted average of these K nearest neighbors.

The formula for the KNN estimation is:

y_hat = Σ(w_i * y_i) / Σ(w_i) (2)

where:

- y_hat is the estimated value for the missing data point
- w_i is the weight assigned to the i-th neighbor (usually based on the inverse of the distance between the missing data point and the i-th neighbor)
- y_i is the value of the i-th neighbor
- The summations (Σ) are taken over the K nearest neighbors

In the context of the Hull strain gauge data of the ARAON, the KNN model was built using the following steps:

1. Feature selection: The relevant independent variables were identified through a correlation analysis, similar to the linear regression model. These variables were used as the features for determining the nearest neighbors.
2. Distance metric selection: A suitable distance metric was chosen to measure the similarity between data points in the feature space. Common distance metrics include Euclidean, Manhattan, and Minkowski distances.
3. Determining the optimal K: Various values of K were evaluated to find the optimal configuration for the KNN model. This was done by assessing the model's performance using metrics such as mean squared error (MSE), mean absolute error (MAE), and R-squared (R2). Cross-validation was also employed to avoid overfitting

and ensure that the chosen K value generalizes well to new data.

4. Missing data estimation: For each missing data point, the K nearest neighbors were identified using the chosen distance metric. The missing value was then estimated by computing the weighted average of the values of these neighbors, as per the formula mentioned above.

KNN has the advantage of being a flexible and easy-to-understand method, but it can be sensitive to the choice of K and the distance metric. Additionally, it may not perform well when dealing with high-dimensional data or in cases where the relationships between variables are complex and non-linear.

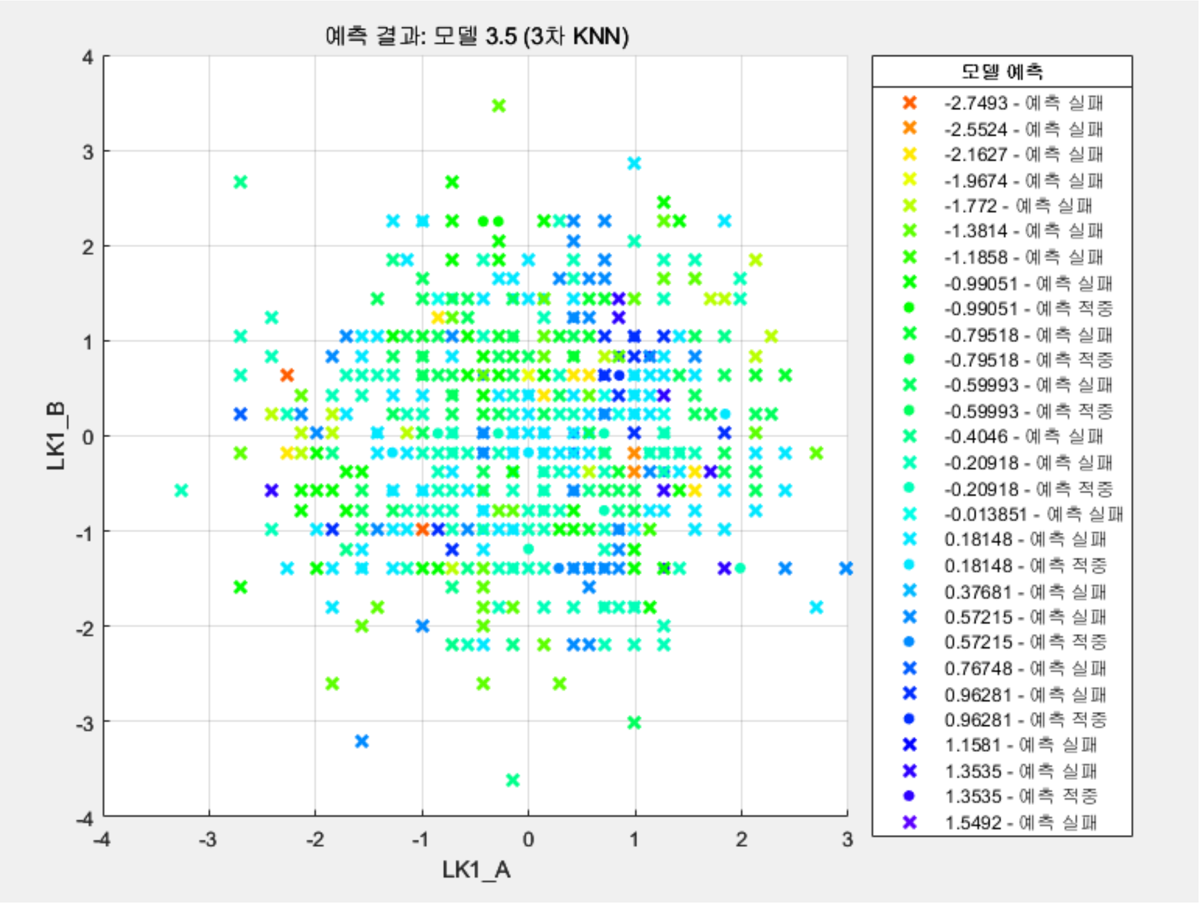


Figure 5. KNN Prediction Model

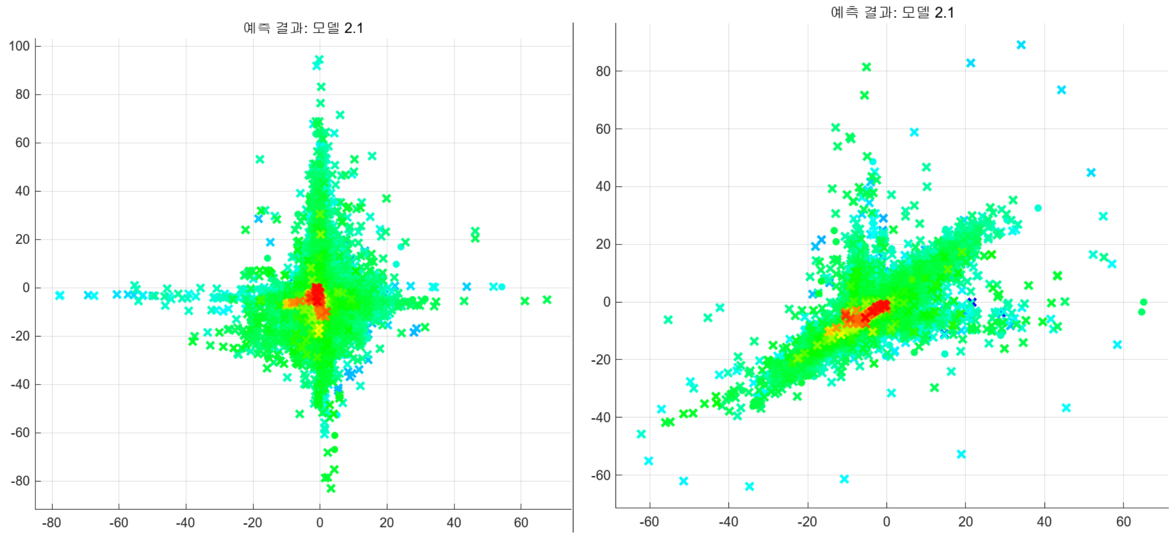


Figure 6. KNN Prediction Model

Our study revealed that the K-Nearest Neighbors (KNN) algorithm performed poorly in estimating the missing Hull strain gauge data of the ARAON, achieving a prediction accuracy of only 10%. Despite exploring several models, the prediction accuracy did not exceed 15%, indicating an overall low performance.

This suboptimal performance can be attributed to a multitude of factors. Firstly, the value of K, representing the number of nearest neighbors in the KNN algorithm, may have been suboptimally chosen. This critical hyperparameter can lead to overfitting or underfitting if inappropriately selected, thereby affecting the model's performance. It is plausible that the K value used in our study failed to accurately capture the local patterns in the data.

Another factor that could have affected the KNN performance is the distance metric used. The choice of distance metric is pivotal in the KNN algorithm, as an inaccurate representation of the relationships between data points can compromise the algorithm's performance. In our study, it is possible that the selected distance metric was not suitable for the Hull strain gauge data.

Furthermore, the KNN algorithm can struggle with high-dimensional data due to the "curse of dimensionality," a phenomenon where the distances between data points become more uniform as the number of dimensions increases. This makes it difficult for the KNN algorithm to distinguish between close and distant neighbors, potentially resulting in poor performance when estimating missing values. In our case, the high dimensionality of the Hull strain gauge data could have negatively affected the KNN's ability to accurately estimate missing values.

The complexity of non-linear relationships in the data could also have contributed to the subpar performance of KNN. As a non-parametric method, KNN does not make any assumptions about the underlying data distribution, and while this can be advantageous in some cases, it may limit the algorithm's ability to capture complex non-linear relationships in the data. It is likely that the Hull strain gauge data exhibited such complex relationships.

Another possible cause for the poor performance could be the missing data mechanism. If the missing data is not missing at random, KNN could produce biased estimates. The missing values in our dataset might have been systematically related to the underlying data, causing KNN to inaccurately estimate them due to its assumption that the observed data points are representative of the entire dataset.

Lastly, scalability issues could have played a part in hindering the performance of the KNN algorithm. KNN can be computationally expensive, particularly for large datasets, as the time and resources required to calculate the distances between the query point and all other data points can be substantial.

3.3.3 Long Short-Term Memory (LSTM) Neural Networks

Long Short-Term Memory (LSTM) neural networks are a type of recurrent neural network (RNN) designed to handle time-series data. LSTMs can learn long-range dependencies and capture temporal patterns, making them well-suited for estimating missing data in time-series data like the Hull strain gauge data of the ARAON.

An LSTM cell contains three main components: an input gate, a forget gate, and an output gate. These gates control the flow of information through the cell, allowing it to learn, remember, and forget information as needed. The LSTM cell also has a hidden state and a cell state that store information over time.

The core equations governing the behavior of an LSTM cell are:

- Input gate (i_t): i_t = σ(W_i * [h_(t-1), x_t] + b_i)
- Forget gate (f_t): f_t = σ(W_f * [h_(t-1), x_t] + b_f)
- Cell state update (C~_t): C~t = tanh(W_C * [h(t-1), x_t] + b_C)
- New cell state (C_t): C_t = f_t * C_(t-1) + i_t * C~_t
- Output gate (o_t): o_t = σ(W_o * [h_(t-1), x_t] + b_o)
- Hidden state (h_t): h_t = o_t * tanh(C_t)

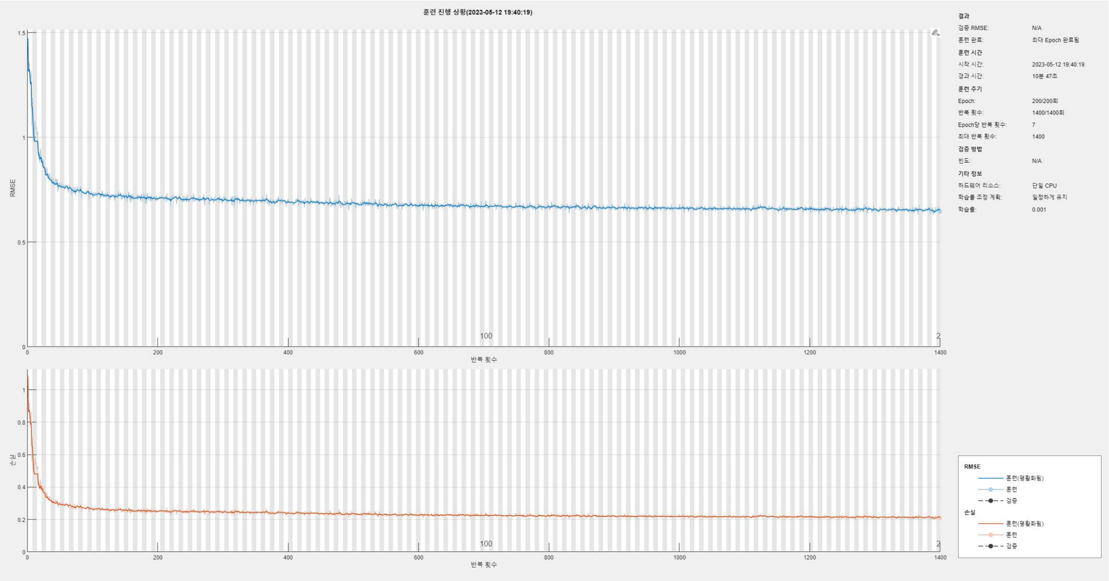
where:

- σ is the sigmoid activation function
- tanh is the hyperbolic tangent activation function
- W and b are the weight matrices and bias vectors for each gate, respectively
- x_t is the input at time step t
- h_(t-1) is the hidden state at the previous time step (t-1)
- C_(t-1) is the cell state at the previous time step (t-1)

In the context of the Hull strain gauge data of the ARAON, the LSTM model was built using the following steps:

1. Data preparation: The time-series data was prepared for the LSTM model by dividing it into sliding windows of fixed size. Each window contained a sequence of strain gauge readings and the corresponding target value (i.e., the next reading in the sequence).

2. Model architecture: The LSTM model's architecture was designed with multiple layers of LSTM cells, followed by dense layers for output estimation. The number of layers, LSTM units, and dense layer units were determined through experimentation and cross-validation to achieve the best performance.
3. Model training: The LSTM model was trained using backpropagation through time (BPTT) and a suitable optimization algorithm (e.g., Adam or RMSprop). The model's performance was monitored using metrics such as mean squared error (MSE) and mean absolute error (MAE).
4. Missing data estimation: The trained LSTM model was used to estimate the missing data points in the strain gauge data. Given a sequence of known readings as input, the model predicted the next reading in the sequence, which could be used as the estimated value for a missing data point



결과

검증 RMSE: N/A

훈련 완료: 최대 Epoch 완료됨

훈련 시간

시작 시간: 2023-05-12 19:40:19

경과 시간: 10분 47초

훈련 주기

Epoch: 200/200회

반복 횟수: 1400/1400회

Epoch당 반복 횟수: 7

최대 반복 횟수: 1400

검증 방법

빈도: N/A

기타 정보

하드웨어 리소스: 단일 CPU

학습률 조정 계획: 일정하게 유지

학습률: 0.001

Figure 7. Training progress of Long Short-Term Memory (LSTM)

In our study, we employed Long Short-Term Memory (LSTM) neural networks to estimate the missing Hull strain gauge data of the ARAON. The LSTM model's performance was evaluated using the Mean Squared Error (MSE) metric. Our results demonstrated that the LSTM model achieved an MSE of approximately 0.5. This value, while not definitive in terms of accuracy, provides insight into the model's performance and can be further analyzed alongside other metrics or in comparison to alternative models.

4 CONCLUSIONS

This study aimed to estimate missing Hull strain gauge data of the ARAON using various artificial intelligence techniques, namely Linear Regression, K-Nearest Neighbors (KNN), Long Short-Term Memory (LSTM) Neural Networks. Each of these techniques demonstrated different strengths and weaknesses when applied to the problem of estimating missing data.

The results revealed that no single technique was universally superior, with each method having its own set of advantages and limitations. Linear Regression offered simplicity and interpretability, KNN excelled in capturing local patterns, LSTM was effective in handling long-range temporal dependencies. Considering the diverse strengths of these methods, the most effective approach for estimating missing Hull strain gauge data of the ARAON may be a combination of these techniques or an ensemble approach. By combining the outputs of multiple models, the overall accuracy and reliability of the estimations can be improved, effectively leveraging the strengths of each method while mitigating their weaknesses.

Further research can explore the development of hybrid models or ensemble techniques to enhance the estimation of missing data points in the Hull strain gauge data. Additionally, researchers could investigate the use of other AI techniques, such as Graph Neural Networks (GNNs) or Autoencoders, to expand the range of potential solutions for this problem.

In conclusion, this study highlights the potential of artificial intelligence techniques for estimating missing data in the Hull strain gauge data of the ARAON. By employing a combination of these methods or developing ensemble approaches, it is possible to increase the accuracy and reliability of the estimations, providing valuable information for ship management and maintenance.

ACKNOWLEDGEMENTS

This research was supported by a grant from Endowment Project of “Development of Evaluation Technology for Ship's Performance in Extreme Environment” funded by the Korea Research Institute of Ships and Ocean Engineering(PES4731)

REFERENCES

Donders, A. R. T., Van Der Heijden, G. J., Stijnen, T., & Moons, K. G. (2006). A gentle introduction to imputation of missing values. Journal of clinical epidemiology, 59(10), 1087-1091.

Lee, J. H., Kwon, Y. H., Rim, C. W., & Lee, T. K. (2016). Characteristics analysis of local ice load signals in ice-covered waters. International Journal of Naval Architecture and Ocean Engineering, 8(1), 66-72.

Liao, T. W. (2005). Clustering of time series data—a survey. Pattern recognition, 38(11), 1857-1874.

Sardá-Espinosa, A. (2017). Comparing time-series clustering algorithms in r using the dtwclust package. R package vignette, 12, 41.

Min, J. K., Cheon, E-J., Kim, J. M., Lee, S. C., & Choi, K. (2016). Comparison of the 6-DOF Motion Sensor and Strain Gauge Data for Ice Load Estimation on IBRV ARAON. Journal of the Society of Naval Architects of Korea, 53(6), 529-535.

Cho, S., Choi, K., Son, B., Jeong, S-Y., & Ha, J-S. (2021). Enhanced Influence Coefficient Matrix for Estimation of Local Ice Load on the IBRV ARAON. Journal of the Society of Naval Architects of Korea, 58(5), 330-338.

Jeon, M., Choi, K., Min, J. K., & Ha, J. S. (2018). Estimation of local ice load by analyzing shear strain data from the IBRV ARAON's 2016 Arctic voyage. International Journal of Naval Architecture and Ocean Engineering, 10(3), 421-425.

Kong, S., Cui, H., Wu, G., & Ji, S. (2021). Full-scale identification of ice load on ship hull by least square support vector machine method. Applied Ocean Research, 106, 102439.

Kim, S., & Kim, D. (2018). Imputation Method for Missing Data Based on Clustering and Measure of Property. The Korean Journal of Applied Statistics, 31(1), 29-40.

Lee, K., Ju, H., & Jeong, Y. M. (2021). Missing Value Prediction System based on AutoML for Sensor maintenance in Edge Networks. The Institute of Electronics and Information Engineers, 1970-1971.

Cho, S., & Choi, K. (2019). Modification of Local Ice Load Prediction Formula Based on IBRV ARAON's Arctic Field Data. Journal of Ocean Engineering and Technology, 33(2), 161-167.

Cini, A., Marisca, I., & Alippi, C. (2021). Multivariate Time Series Imputation by Graph Neural Networks. arXiv preprint arXiv:2108.00298.